



UNIVERSITÄT ZU LÜBECK  
INSTITUT FÜR MULTIMEDIALE  
UND INTERAKTIVE SYSTEME

# **Entwicklung eines Interfaces für ein Entscheidungsunterstützungssystem für die Planung von Lebensmittelkontrollen**

Development of an interface for a decision support system  
for the planning of food controls

**Bachelorarbeit**

im Rahmen des Studiengangs

**Medieninformatik**

der Universität zu Lübeck

vorgelegt von:

**Ilja Jamkis**

ausgegeben und betreut von:

**Prof. Dr. rer. pol. Moreen Heine**

mit Unterstützung von:

**Pascal Stieglitz M.Sc.**

Kiel, 3. Dezember 2023



# Kurzfassung

Die behördliche Lebensmittelüberwachung leistet einen wichtigen Teil zur Erhalt der öffentlichen Sicherheit. Durch zunehmende Anforderungen bei gleichzeitig steigendem Personalmangel ist in diesem Bereich der Bedarf für KI-basierte Unterstützungssysteme hoch. In dieser Arbeit wurde ein Interface für ein KI-basiertes Entscheidungsunterstützungssystem entworfen, das Lebensmittelkontrolleuren helfen soll, zu entscheiden, welche Betriebe priorisiert kontrolliert werden müssen.

Das entwickelte System soll die Nutzer dabei unterstützen, die Berechnung der Risikobewertung und dahinterliegenden Prozesse nachvollziehen zu können. Zu diesem Zweck stellt das System Erklärungen und Begründungen in verschiedenen Formaten bereit, die die Entscheidungsfindung des KI-Systems und die beteiligten Prozesse transparent machen sollen.

Da die Entscheidung, welche Betriebe kontrolliert werden, einen Entscheidungsspielraum umfasst, darf sie nicht allein von einem KI-System getroffen werden. Darum ist es zentral, dass den Kontrolleuren die Gründe, warum ein Betrieb kontrolliert werden sollte, in einer Form präsentiert werden, die für sie nachvollziehbar ist. Daher ist eine nutzerzentrierte Gestaltung des Systems wichtig.

Für die Umsetzung dieses Ziels wurden zwei Bereiche betrachtet. Zum einen wurde der Forschungsbereich Explainable AI betrachtet, der sich damit beschäftigt, wie Nutzern die Ergebnisse und Prozesse eines KI-Systems verständlich erklärt werden können. Zum anderen wurde der Nutzungskontext der Lebensmittelüberwachung analysiert. Daraus wurden Anforderungen an das System abgeleitet. In einem iterativen Gestaltungsprozess wurde ein High-Fidelity-Prototyp entwickelt.

Um zu überprüfen, ob die Ziele der Arbeit erfüllt wurden, wurde in einer abschließenden Evaluation folgende Frage beantwortet: „Ist ein System, das auf Gestaltungsprinzipien für Explainable AI entwickelt wurde, nachvollziehbarer als ein System, das nicht auf solchen Prinzipien basiert?“. In einer abschließenden Evaluation wurde die Frage mit einem A/B-Test und weiteren Methoden überprüft.

Die Auswertung der Daten lässt darauf schließen, dass die für die Evaluation gestellte Frage, ob ein auf XAI-basiertes System die Ergebnisse eines KI-Systems nachvollziehbarer macht als ein System ohne XAI-Feature, positiv beantwortet werden kann. Sowohl in Bezug auf die subjektive Wahrnehmung als auch objektive Maße (Bearbeitungsdauer) kann XAI eine Unterstützung sein.

Schlüsselwörter: XAI, Explainable AI, Nachvollziehbarkeit, User Interface, Lebensmittelüberwachung

# Abstract

Official food monitoring plays an important role in maintaining public safety. Due to high requirements combined with a growing shortage of staff, the need for AI-based support systems in this field is increasing. In this thesis, an interface for an AI-based decision support system was designed to help food inspectors decide which establishments should be prioritised for inspection.

The developed system should support the users in understanding the calculation of the risk assessment and the underlying processes. To this end, the system provides explanations and justifications in various formats to make the AI system's decision-making process and the processes involved transparent.

The decision which locations need to be inspected cannot not be made by an AI system alone. It is therefore crucial that the reasons why a site should be inspected are presented to the inspectors in a form that they can understand. A user-centred design of the system is therefore important.

Two approaches were followed for the realisation of this goal. Firstly, the research area of Explainable AI was analysed, which deals with how the results and processes of an AI system can be explained to users in an understandable way. Secondly, the usage context of food monitoring was analysed. Requirements for the system were derived from this. A high-fidelity prototype was developed in an iterative design process.

In order to check whether the objectives of the work were met, the following question was answered in a final evaluation: "Is a system which was developed based on design principles following Explainable AI more comprehensible than a system that is not based on such principles?". In a final evaluation, the question was tested using an A/B test and other methods.

The analysis of the data suggests that the question posed for the evaluation if a system based on XAI makes the results of an AI system more comprehensible than a system without an XAI feature can be answered with yes. XAI can provide support in terms of both subjective perception and objective measures.

**Key Words:** XAI, Explainable AI, Traceability, User Interface, Food Control

## Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Ziele der Arbeit.....	2
1.2	Vorgehensweise.....	3
2	Theoretische Hintergründe zu KI-Systemen in der öffentlichen Verwaltung.....	5
2.1	Einführung in Entscheidungsunterstützungssysteme in Behörden .....	5
2.2	Akzeptanz von Entscheidungsunterstützungssystemen in Behörden.....	6
2.3	Die „Ethik-Leitlinien für eine vertrauenswürdige KI“ der EU und ihre Empfehlungen zum Einsatz von automatisierten Entscheidungen in Behörden.....	7
2.4	Explainable AI (XAI).....	8
2.5	Fazit.....	13
3	Analyse .....	14
3.1	Literaturrecherche.....	14
3.2	Faktoren für die Gestaltung von XAI-Systemen.....	15
3.3	Datenquellen der Nutzungskontextanalyse .....	21
3.4	Nutzungskontextanalyse.....	21
3.4.1	Die Aufgaben der Lebensmittelkontrolleure.....	22
3.4.2	Die Nutzer .....	25
3.5	Fazit der Analyse .....	26
4	Konzeption .....	28
4.1	Analyseworkshop.....	28
4.2	Use-Case .....	30
4.3	Funktionalitäten .....	32
4.4	Annahmen für das KI-System und Interface .....	34
4.5	Interfacedesign.....	35
4.5.1	1. Iteration: Skizzen .....	35
4.5.2	2. Iteration: Mockups .....	41
4.6	Systemarchitektur.....	49
4.7	Fazit.....	51
5	Realisierung .....	52
5.1	Verwendete Werkzeuge .....	52
5.3	Struktur des Projekts.....	54
5.4	Implementierung der Funktionalitäten .....	55
5.4.1	Umsetzung der Content Verwaltung.....	55
5.4.2	Umsetzung des Interfaces .....	58
5.5	Fazit.....	63
6	Dialogbeispiele.....	64

7 Summative Evaluation.....	68
7.1 Ziel.....	68
7.2 Methode.....	69
7.2.1 Design.....	69
7.2.2 Teilnehmer.....	70
7.2.3 Setting und Instrumente.....	71
7.2.4 Ablauf.....	75
7.3 Ergebnisse.....	77
7.3.1 Wahrgenommene Nachvollziehbarkeit (SIPA-Skala).....	77
7.3.2 Beanspruchung durch Bearbeitung der Aufgaben (Task Load Index).....	79
7.3.3 Ergebnisse der Zeitmessung.....	83
7.3.4 Ergebnisse zur Gebrauchstauglichkeit (SUS-Werte).....	86
7.3.5 Ergebnisse aus dem Quiz – Aufgabe 1 „Verständnis über das System“.....	87
7.4 Diskussion.....	88
7.4.1 Vergleich der SIPA-Mittelwerte.....	88
7.4.2 Vergleich der Nutzerbeanspruchung durch die Bearbeitung der Aufgaben.....	88
7.4.3 Zusammenhang der Bearbeitungsdauer mit der Reihenfolge der Systemvariantennutzung.....	90
7.4.4 Vergleich der Ergebnisse zur Gebrauchstauglichkeit.....	91
7.4.5 Limitierung.....	92
7.5 Fazit Evaluation.....	92
8 Fazit.....	94
8.1 Zusammenfassung.....	94
8.2 Offene Punkte.....	95
8.3 Ausblick.....	96
Abbildungsverzeichnis.....	98
Tabellenverzeichnis.....	99
Literaturverzeichnis.....	100
Anhang.....	103
Anhang A: Interviewleitfaden KI-DemÖ Mitarbeiter.....	103
Anhang B: Interviewleitfaden Lebensmittelkontrolleur.....	106
Anhang C: LimeSurvey Umfrage.....	109
Anhang D: Beispieldatensatz.....	121
Anhang E: Erklärung.....	132

**Hinweis zur Verwendung des generischen Maskulinum:**

Zur besseren Lesbarkeit wird in dieser Hausarbeit das generische Maskulinum verwendet. Die in dieser Arbeit verwendeten Personenbezeichnungen beziehen sich – sofern nicht anders kenntlich gemacht – auf alle Geschlechter.

# 1 Einleitung

Die Lebensmittelkontrolle leidet wie viele Bereiche der öffentlichen Verwaltung unter einem sich immer weiter zuspitzenden Fachkräftemangel. Foodwatch kritisierte schon 2020, dass 90 Prozent der Behörden in Deutschland nicht ausreichend kontrollieren (Born, 2020). In weiten Teilen des Bundesgebiets weichen die Zahlen der gemachten Kontrollen von den gesetzlich vorgeschriebenen Pflichtkontrollen ab. Dies liegt häufig an der personellen Unterbesetzung (Foodwatch, 2019). Gleichzeitig ist der Bereich der Lebensmittelkontrolle unverzichtbar für die Sicherung der öffentlichen Gesundheit. Immer wieder kommt es zu Lebensmittelskandalen mit teils tödlichen Folgen (Haunhorst, 2019).

Doch auch mit dem dringend benötigten zusätzlichen Personal ist es für die Lebensmittelkontrolle unrealistisch, alle Betriebe lückenlos zu kontrollieren. In welchen Abständen Kontrollen stattfinden, ist in der Verordnung Allgemeine Verwaltungsvorschrift über Grundsätze zur Durchführung der amtlichen Überwachung [...] (kurz AVV RÜb) definiert. Einerseits basiert die Auswahl auf einer Risikoanalyse, die Faktoren wie Art des Betriebs, Produktart oder Verhalten des Unternehmers beinhaltet. Andererseits werden Kontrolleure auch nach Beschwerden aus der Bevölkerung oder auf Verdacht hin aktiv (Verbraucherbeschwerde, o. J.).

Eine Künstliche Intelligenz (KI) als Entscheidungsunterstützungssystem kann bei der Auswahl der zu kontrollierenden Betriebe helfen, indem sie große Datenmengen aus verschiedenen Quellen zusammenträgt, miteinander verknüpft, auswertet und in einem für die User gebrauchstauglichen Format zur Verfügung stellt.

Auch können soziale Netzwerke und Rezensionsportale im Internet durch ein KI-System nach relevanten Auffälligkeiten durchsucht werden (von Lucke & Etscheid, 2020). Der Einsatz eines KI-Systems kann durch Erhöhung der Trefferquote zur Folge haben, dass der Verbraucherschutz im Bereich Lebensmittelsicherheit verbessert wird. Die Zahl der durchgeführten Inspektionen bleibt vergleichbar mit dem Szenario ohne KI. Insgesamt steigt aber damit die Wahrscheinlichkeit, Verstöße aufzudecken und die Bevölkerung erfolgreich zu schützen.

Ein positives Zusammenspiel zwischen KI-System und User ist jedoch nicht selbstverständlich. Damit aus der Interaktion ein Mehrwert geschaffen wird, gilt es den User und dessen Entlastung im Fokus zu behalten (von Lucke & Etscheid, 2020).

Ein Aspekt, der im Kontext von KI-Anwendung und Lebensmittelkontrolle besonders interessant scheint, ist die Nachvollziehbarkeit von KI-Entscheidungen. Da bei einem Entscheidungsunterstützungssystem die endgültige Entscheidung, ob ein Betrieb kontrolliert wird oder nicht, beim Menschen liegt und dieser damit auch die Verantwortung trägt, muss das von dem KI-System erarbeitete Ergebnis für den User nachvollziehbar sein. Die



Verwaltungsmitarbeiter können nur dann sinnvoll mit einer Empfehlung arbeiten, wenn transparent ist, wie sie zustande gekommen ist. Außerdem muss die Empfehlung überprüfbar sein (von Lucke & Etscheid, 2020).

Kann ein User die Entscheidungen eines Systems nicht nachvollziehen, kann dies dazu führen, dass er das System nicht nutzt oder die ausgegebenen Entscheidungsvorschläge nicht angemessen kritisch hinterfragt. Aktuelle Studien der Bertelsmann Stiftung zeigen, dass ohnehin Misstrauen gegenüber KI-Systemen herrscht. So lehnen mehr als drei Viertel der Befragten in Deutschland vollautomatisierte Entscheidungen ab (Fischer & Petersen, 2018). Diese Umfrage ist Teil eines intensiv geführten Diskurses, sowohl in Wissenschaft als auch Öffentlichkeit, über die Verwendung von künstlicher Intelligenz. So warnen Organisationen wie „algorithmWatch“ davor, dass Menschen durch automatisierte Entscheidungssysteme „in ihrer Teilhabe beeinträchtigt werden, dass ihre Grundrechte beschnitten werden oder dass ihnen der Zugang zu öffentlichen Gütern und Dienstleistungen verwehrt wird“ (Automatisierte Entscheidungssysteme im öffentlichen Sektor, 2022). Auf der anderen Seite stehen die enormen Potenziale von KI im Bereich der öffentlichen Verwaltung (Bernnat et al., 2022).

Der Diskurs ist wichtig, seine Pole sollten aber nicht als unvereinbar missverstanden werden. Ziel muss es sein, KI-Systeme zu demokratisieren und verstehbar zu machen und somit anwendbar in sensiblen Bereichen. Es braucht das Verständnis und die Akzeptanz der Nutzer (und zu einem bestimmten Maß auch der breiten Öffentlichkeit). Diese Ansprüche müssen sich auch bei der Nutzung von KI-Anwendungen im Bereich Lebensmittelkontrolle widerspiegeln. Wie das konkret gelingen kann, soll Thema dieser Arbeit sein.

## 1.1 Ziele der Arbeit

Im Rahmen dieser Bachelorarbeit wird für den thematischen Kontext der Lebensmittelüberwachung ein Interface entwickelt, das für die Aufgabe der Planung von Lebensmittelkontrollen eine Auswahl von zu kontrollierenden Betrieben vorschlägt und diese erklärt. Das Interface ist dabei die Schnittstelle zwischen einem fiktiven KI-basierten Entscheidungsunterstützungssystem und den menschlichen Nutzern.

Primäres Ziel dieser Arbeit ist es, die Ergebnisse eines fiktiven KI-Systems übersichtlich und nachvollziehbar darzustellen. Dazu wird sich an den Methoden und Ideen des Explainable Artificial Intelligence (XAI) Ansatzes orientiert. Das sekundäre Ziel dieser Arbeit ist, den Prozess der Planung von Lebensmittelkontrollen zu verstehen und aufbauend auf diesem Verständnis eine Gestaltungslösung zu entwickeln, die die Nutzer bei der Aufgabe unterstützen. Beide Ziele werden über die Entwicklung eines Interfaces erreicht, welches als Webanwendung implementiert wird.

Dabei wird die Vielzahl an Daten durch verständliche Kennzahlen oder Visualisierungen leicht verständlich dargestellt. Da das KI-basierte Entscheidungsunterstützungssystem fiktiv ist, wird

für diese Arbeit ein System definiert, das verschiedene Datenquellen und für den Nutzungskontext wichtige Parameter nutzt. Dieses soll über eine geeignete Schnittstelle dem zu implementierenden Interface eine Mindestzahl von benötigten Betrieben bereitstellen.

Zusammengefasst sind die Ziele der Arbeit das Verstehen des Nutzungskontextes Lebensmittelüberwachung, das Anknüpfen an die Forschung zu Nachvollziehbarkeit von KI und, auf Basis dieses Wissens, das Entwickeln eines Systems, das die KI-Ergebnisse den Lebensmittelkontrolleuren nachvollziehbar und übersichtlich präsentiert.

## 1.2 Vorgehensweise

Die Vorgehensweise folgt grundsätzlich dem menschenzentrierten Gestaltungsprozess nach (DIN EN ISO 9241-210) und ist daher in Analyse-, Konzeptions-, Realisierungs- und Evaluationsphase unterteilt.

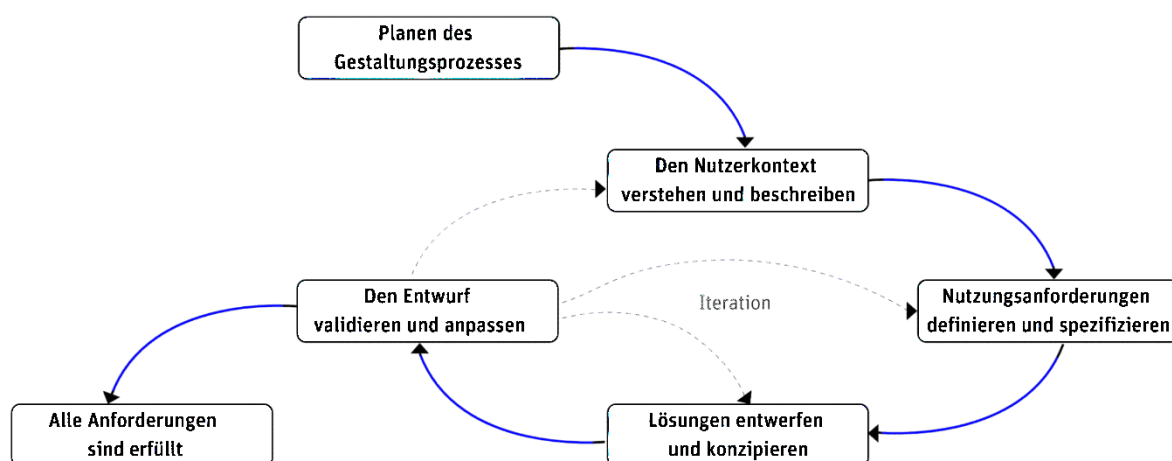


Abbildung 1: Menschenzentrierter Gestaltungsprozess (DIN EN ISO 9241-210)

Zu Beginn des menschenzentrierten Gestaltungsprozesses wird das Erlangen eines möglichst tiefgehenden Verständnisses des Nutzungskontextes angestrebt. Schnell wurde bei der Beschäftigung mit dem Kontext deutlich, dass in der deutschen öffentlichen Verwaltung besondere Anforderungen an KI-Systeme gestellt werden. Deswegen wird ab Kapitel 2.1 zunächst der Status quo der Einführung von Entscheidungsunterstützungssystemen in der öffentlichen Verwaltung beleuchtet, sowie die rechtliche Lage in Deutschland und in der EU als wichtige Grundlage für die Akzeptanz von Entscheidungsunterstützungssystemen beschrieben. Anschließend wird sich in Kapitel 2.4 der Anforderung der Nachvollziehbarkeit von KI-Systemen genähert und damit einhergehend der Forschungsbereich XAI vorgestellt.

In Kapitel 3 folgt die Analyse und das Aufstellen der Nutzungsanforderungen. Ermittelt werden Faktoren bzw. Kriterien für Nachvollziehbarkeit von KI-Prozessen und daraus abgeleitete Designempfehlungen für die Mensch-KI-Interaktion (Kapitel 3.2). Danach wird in Kapitel 3.4 der Nutzungskontext der Lebensmittelüberwachung und die damit verbundenen Nutzer und seine Aufgaben eingehend beschrieben.

Aufbauend auf den Ergebnissen der Analyse wurde in der Konzeptionsphase ein Gestaltungskonzept entwickelt (Kapitel 4). Den gesammelten Anforderungen wurden Funktionalitäten zugeschrieben und es wurde eine Priorisierung dieser Anforderungen bzw. Funktionen vorgenommen (Kapitel 4.3). Anschließend wurde in zwei Iterationen das Interface Design des entstehenden Systems festgelegt (Kapitel 4.5). Da in dieser Arbeit lediglich ein Interface-Prototyp entwickelt wurde, wird in Kapitel 4.6 die Systemarchitektur eines Produktivsystems vorgestellt, um die verschiedenen Komponenten vorzustellen, die an so einem System beteiligt sind.

Der Konzeption folgte die Realisierung des Systems als High-Fidelity-Prototyp (Kapitel 5).

In Kapitel 6 wird ein Dialogbeispiel vorgestellt.

Abschließend wurde die Evaluation des implementierten Systems durchgeführt (Kapitel 7). Die Evaluation erfolgte durch einen A/B-Test inkl. Interaktion mit dem fertig implementierten System. Es sollte die Frage beantwortet werden, ob ein System, das auf XAI-Prinzipien basiert, nachvollziehbarer ist als ein System, das nicht darauf basiert.

## 2 Theoretische Hintergründe zu KI-Systemen in der öffentlichen Verwaltung

Im ersten Teil dieses Kapitels soll der aktuelle Stand zur Einführung von KI-Systemen in der deutschen öffentlichen Verwaltung dargestellt werden. Auch sollen ausgewählte, einhergehende Diskussionen beleuchtet werden. Das Betrachten des Status quo ist notwendig, um die Herausforderungen bei der Einführung von KI-gestützten Systemen zu verstehen. Im zweiten Teil des Kapitels wird der Forschungsbereich XAI vorgestellt.

### 2.1 Einführung in Entscheidungsunterstützungssysteme in Behörden

KI-Systeme können in der öffentlichen Verwaltung für die Erledigung vieler verschiedener Aufgaben eingesetzt werden. Das Fraunhofer Institut stellt in seiner Studie „Künstliche Intelligenz in der öffentlichen Verwaltung“ von 2020 verschiedene, beispielhafte Einsatzgebiete von KI-Systemen in Behörden vor und definiert Typen von KI-gestützten Entscheidungsunterstützungssystemen (Etscheid et al., 2020).

Viele digitale, KI-gestützte Anwendungen in der öffentlichen Verwaltung finden im sog. Front-Office, also im Bürgerkontakt, etwa durch Chat-Bots statt. Auch in der Hintergrundverwaltung, im Back-Office werden KI-gestützte Anwendungen bereits erfolgreich eingesetzt (Etscheid et al., 2020). Beratende Systeme sind der KI-Typ, der in der deutschen Verwaltung am häufigsten zur Anwendung kommt und in dem der Grad der automatisierten Entscheidungsunterstützung am stärksten vorangeschritten ist. Die Verwaltung ist ein Bereich, der traditionell stark von Gesetzen, Normen und Vorschriften geprägt ist. Grundlegend für Verwaltungshandeln (ausgeführt durch die Verwaltungsbeschäftigten) ist, Rechtsvorschriften korrekt auf einen Einzelfall zu beziehen (Martini & Nink, 2018). Oft ist jedoch trotz Rechtsvorschrift eine Entscheidung zwischen verschiedenen Alternativen von Nöten. Manche Entscheidungsfindungen sind anhand von Vorschriften einfach, für andere müssten eine Vielzahl von Daten und komplexer Zusammenhänge ausgewertet werden. In der Realität müssen Verwaltungsbeschäftigte in sehr kurzer Zeit sehr viele Entscheidungen treffen oder haben nur einen limitierten Zugang zu Informationen und können daher diese Auswertungen und Abwägungen oft nur oberflächlich treffen (Etscheid et al., 2020).

Neben dieser gebundenen Entscheidungsebene gibt es jedoch auch noch Rechtsnormen, die der Behörde bzw. den Beschäftigten Ermessensspielraum ermöglichen bzw. abverlangen. Dies bedeutet, dass Kriterien abgewogen werden müssen und mehrere Entscheidungen möglich und rechters sind. Hier braucht es einen befugten Menschen, der im Auftrag der Verwaltung diese Entscheidung treffen muss. Beratende Systeme erkennen dies an und treten in eine Interaktion mit dem menschlichen Anwender. Sie können relevante Daten auswerten und einfach

verständlich, z.B. durch Visualisierungen, darstellen (Etscheid et al., 2020). Beratende Systeme unterstützen so den Anwender, die bestmögliche Entscheidung zu treffen. Außerdem lernen sie aus den Entscheidungen des Users und können diese zukünftig berücksichtigen. Beispiele sind die intelligente Einsatzplanung oder vorausschauende Wartung (Etscheid et al., 2020).

Wichtig ist bei beratenden Systemen, dass bei der Erstellung der KI sowie im Interface-Design stets deutlich wird, dass der menschliche Nutzer die Entscheidung trifft und nicht das KI-Entscheidungsunterstützungssystem.

Im September 2017 wurde das Verwaltungsverfahrensgesetz (VwVfG) um den Paragraphen §35a ergänzt, der vollständig automatisierte Erlasse von Verwaltungsakten zulässt. Dies jedoch nur, „sofern dies durch eine Rechtsvorschrift zugelassen ist und weder ein Ermessen noch ein Beurteilungsspielraum besteht“ (Vollständig automatisierter Erlass eines Verwaltungsaktes, 2017).

Ermessensentscheidungen sind also derzeit noch von vollständiger Automatisierung der Entscheidung ausgeschlossen. Sie sollen ausschließlich beim Menschen liegen.

## 2.2 Akzeptanz von Entscheidungsunterstützungssystemen in Behörden

Der Erfolg der Einführung eines neuen IT-Systems hängt auch davon ab, ob seine zukünftigen Nutzer es akzeptieren (Engelmann & Puntschuh, 2020). Gerade bei neuartigen, KI-gestützten Entscheidungsunterstützungssystemen ist dieser menschliche Faktor relevant. Die Verwaltungsbeschäftigten müssen mit den Systemen umgehen können und eine Erleichterung in ihrem Arbeitsalltag spüren (Engelmann & Puntschuh, 2020).

Aus diesem Grund sollte Akzeptanz von Entscheidungsunterstützungssystemen in Behörden bei der Analyse des Nutzungskontextes unbedingt mitgedacht werden. Die zukünftigen Anwender sowie ihre Personalvertretungen und Vorgesetzten sollten von Anfang an, und anschließend regelmäßig, in den Prozess der Konzeption eingebunden werden, um ein System zu entwickeln, dass den Bedürfnissen der Nutzer entspricht. Um einen Eindruck zu gewinnen, inwieweit KI-gestützte Entscheidungsunterstützungssysteme in der deutschen öffentlichen Verwaltung akzeptiert sind, sollten drei Ebenen betrachtet werden. Der rechtlich-organisatorische Rahmen, die Beschäftigten und die Bevölkerung als Nutzer der staatlichen Angebote.

Die rechtliche Situation zu KI-Systemen in der Verwaltung wurde im vorherigen Kapitel beschrieben.

Soll die Akzeptanz von KI-Systemen bei Mitarbeitern in der deutschen Verwaltung analysiert werden, fällt auf, dass zu dieser Fragestellung wenige wissenschaftliche Veröffentlichungen vorhanden sind. Aus diesem Grund wird im Folgenden zunächst auf Studien verwiesen, die die Akzeptanz von KI-Systemen innerhalb der gesamten deutschen Bevölkerung untersucht haben.

Die Ergebnisse einer repräsentativen Umfrage der Bertelsmann Stiftung von 2022 (Overdiek & Petersen, 2022) deuten darauf hin, dass viele Menschen in Deutschland noch Vorbehalte gegenüber KI haben. Allerdings ändert sich das Bild. Da die Umfrage 2018 und 2021 durchgeführt wurde, wird deutlich, dass Unwissen und Misstrauen gegenüber KI-Systemen und Algorithmen abnimmt (Overdiek & Petersen, 2022). Präferierten 2018 noch 79 % der Befragten menschliche Entscheidungen gegenüber automatisierten (Fischer & Petersen, 2018), zeigen die Ergebnisse von 2022, dass die Akzeptanz von KI-gestützten Anwendungen „spürbar gestiegen“ (Overdiek & Petersen, 2022) ist.

Auch wenn seit 2018 das Wissen über KI in der Bevölkerung gestiegen ist, sollte bei der Konzeption von KI-gestützten Verfahren immer noch mit großen Wissenslücken und Vorbehalten gerechnet werden. Bei der Entwicklung ist es wichtig, zielgruppenorientiert und möglichst niedrigschwellig zu gestalten.

## 2.3 Die „Ethik-Leitlinien für eine vertrauenswürdige KI“ der EU und ihre Empfehlungen zum Einsatz von automatisierten Entscheidungen in Behörden

Um bei der Entwicklung eines KI-Systems den Ansprüchen von Behörden nach vertrauenswürdiger KI gerecht zu werden, ist es wichtig, sich neben der gesetzlichen Lage in Deutschland auch mit internationalen bzw. europäischen Standards auseinanderzusetzen.

Viele internationale Organisationen, wie die OECD, die Vereinten Nationen oder Staaten wie die USA haben Leitlinien oder Gesetze zur Verwendung von KI definiert. Eine zentrale Leitlinie in der EU ist die „Ethik Guideline für vertrauenswürdige KI“, die 2019 veröffentlicht wurde und von einer Expertengruppe für künstliche Intelligenz der Europäischen Kommission entwickelt wurde (Ethics Guidelines for Trustworthy AI, 2019).

Die EU ist eine supranationale Struktur, deren Entscheidungen für die in der EU lebenden Menschen viele Auswirkungen hat. Werden die Leitlinien in EU-Recht überführt, sind die deutschen Behörden verpflichtet, diese umzusetzen bzw. sich an dieses Recht anzupassen.

Die Ethik-Leitlinien für eine vertrauenswürdige KI betonen die Bedeutung von Transparenz und Erklärbarkeit von KI-Systemen. Sie fordern, dass KI-Systeme entwickelt werden sollten, die erklärbare Ergebnisse liefern und den Benutzern die Nachvollziehbarkeit der zugrunde liegenden Entscheidungsprozesse ermöglichen.

Die Leitlinie spricht darüber, wie vertrauenswürdige KI-Systeme verwirklicht werden können. Dafür nennt sie sieben Anforderungen an KI-Systeme: (1) Vorrang menschlichen Handelns und menschliche Aufsicht, 2) technische Robustheit und Sicherheit, 3) Schutz der Privatsphäre und Datenqualitätsmanagement, 4) Transparenz, 5) Vielfalt, Gleichbehandlung und Fairness, 6)

gesellschaftliches und ökologisches Wohlergehen sowie 7) Rechenschaftspflicht (Ethics Guidelines for Trustworthy AI, 2019).

Die Leitlinien schlagen verschiedene Methoden für mehr Erklärbarkeit in KI-Systemen vor. Besonders hervorgehoben wird die Forschung an XAI. Auf XAI wird im folgenden Kapitel genauer eingegangen, da es ein Ansatz ist, der reich an Erkenntnissen und Hinweisen zum Thema Nachvollziehbarkeit ist. Die Anforderungen, und wie sie in der Konzeption von Mensch-KI-Schnittstellen konkret umgesetzt werden können, werden in Kapitel 3.2 genauer beschrieben.

## 2.4 Explainable AI (XAI)

Viele der typischen KI-Systeme gelten als Blackbox-Systeme, da ihre inneren Mechanismen nur schwer darstellbar sind. Diese Undurchsichtigkeit hat den Bedarf an der Entwicklung interpretierbarer KI-Systeme geweckt. Ein ganzer Forschungsbereich namens Explainable Artificial Intelligence (im Folgenden abgekürzt zu XAI) hat sich zum Ziel gesetzt, KI-Systeme erklärbar zu machen (Kamath & Liu, 2021).

Arrieta et al. (2020) nach, kann XAI wie folgt definiert werden:

“Given an audience, an XAI is one that produces details or reasons [regarding a ML model] to make its functioning clear or easy to understand.”

XAI hat seinen Ursprung im Erklärbarmachen der technischen Prozesse von KI-Systemen, wie etwa neuronale Netze oder Machine Learning. Entwickler solcher Systeme haben damit versucht, die inneren Mechanismen ihrer eigenen KI-Modelle erklärbar zu machen (Liao et al., 2020).

### **Pluralistisches Konzept XAI**

Innerhalb von XAI findet zunehmend interdisziplinäre Forschung zu menschenzentrierter XAI statt, die den menschlichen Faktor bei der Erklärung von KI-Systemen in den Fokus nimmt (Lopes et al., 2022; Hoffman et al., 2023).

Durch die zunehmende Menschenzentrierung in XAI wird das, was unter XAI subsummiert wird, immer pluralistischer und komplexer. Nicht nur aufgrund des schwer festzumachenden Geltungsbereichs ist eine eindeutige Definition zurzeit nicht möglich. Auch der Begriff für das Konzept XAI selbst wird noch nicht einheitlich verwendet. Laato et al. (2022), die eine Literaturreview durchgeführt haben, haben verschiedene Begriffe entdeckt, die in der XAI-Forschung verwendet wurden, um das Konzept zu beschreiben. Dazu gehören KI-Interpretierbarkeit, Transparenz und Verständlichkeit. Eine Befragung von Interessenvertretern und Praktikern zeigte, dass bis zu 20 Synonyme für XAI verwendet werden, was den Mangel an einheitlicher Terminologie unterstreicht (Laato et al., 2022)

Der Begriff der Verständlichkeit („understandability“) wird mit mindestens drei weiteren Begriffen synonym verwendet (Laato et al., 2022):

1. interpretability (Interpretierbarkeit - Der Grad, zu welchem ein Beobachter die Gründe für eine Entscheidung nachvollziehen kann (Miller, 2018))
2. intelligibility (dt. „Verständlichkeit“)
3. comprehensibility (dt. „Verständlichkeit“)

Auch der Begriff „erklären“ (explaining) wird häufig synonym mit „understanding“ und „interpreting“ im XAI-Kontext verwendet (Holzinger, 2021; Miller, 2018).

Darüber hinaus wird argumentiert, dass eine Unterscheidung gemacht werden müsse zwischen „explainability“ und „causality“, da „causality“ eine Eigenschaft einer Person und „explainability“ eine Eigenschaft eines Systems ist (Holzinger et al., 2019).

In der Literatur wird diese Uneinheitlichkeit der Begriffe im Konzept des XAI nicht unbedingt als Schwäche beschrieben. Ehsan & Riedl (2022) bestätigen, dass es keinen Konsens darüber gibt, was mit „Erklärbarkeit“ gemeint ist. Sie stellen jedoch die Frage, ob eine allgemeingültige Definition zum jetzigen Zeitpunkt sinnvoll ist und merken an, dass dies weder machbar noch wünschenswert sei.

Obwohl sich die Definitionen und Interpretationen leicht unterscheiden, beziehen sich diese Begriffe alle darauf, wie gut sich die Nutzer die Funktionsweise eines Systems und seine Entscheidungen vorstellen können (Miller, 2018).

### **Dimensionen (und Prinzipien) menschenzentrierter Erklärungen in XAI**

Erklärungen richten sich stets an Menschen, daher scheint es zunächst verwunderlich, wenn explizit von menschenzentrierten Erklärungen gesprochen wird. In manchen Fällen können Erklärungen jedoch technikzentriert sein, was bedeutet, dass sie mehr auf die technischen Details und die Funktionsweise des Systems ausgerichtet sind (wie oben beschrieben), ohne notwendigerweise auf die Bedürfnisse und das Verständnis der Nutzer einzugehen. Menschenzentrierte Erklärungen in der XAI hingegen legen einen besonderen Fokus darauf, dass die Erklärungen für Nicht-Experten verständlich, relevant und nutzbar sind. Sie berücksichtigen psychologische und sozialwissenschaftliche Erkenntnisse darüber, wie Menschen Informationen verarbeiten und verstehen.

Das Hauptziel von menschenzentrierter XAI besteht darin, Nutzer in die Lage zu versetzen, die Mechanismen und Faktoren zu verstehen, wie und warum ein KI-System zu einem bestimmten Ergebnis gekommen ist. Erklärungen können über verschiedene Konzepte/Begriffe beschrieben werden.

Miller (2018) kommt in seiner Veröffentlichung zu vier grundlegenden Erkenntnissen zu Erklärungen:

1. Erklärungen sind kontrastiv. Sie werden als Reaktion auf bestimmte kontrafaktische Fälle gesucht. Das heißt, Menschen fragen nicht, warum Ereignis P passiert ist, sondern vielmehr, warum Ereignis P statt eines Ereignisses Q passiert ist.



2. Erklärungen sind selektiv. Menschen erwarten nicht alle Dimensionen der Wahrheit in der Erklärung eines Vorkommnisses. Sie stellen sich meist mit einem oder zwei Gründen zufrieden. Für die meisten KI-Systeme ist es nicht vorteilhaft, eine Aneinanderreihung von Gründen als Antworten auf „Warum-Fragen“ auszugeben, egal wie sie präsentiert werden, da für Menschen schlicht oft zu komplex. Außerdem ist das Risiko groß, dass die Teile der Erklärungskette, die für die Frage wirklich relevant sind, in der Vielfalt der Gründe untergehen.
3. Obwohl Wahrheit und Wahrscheinlichkeit wichtig für das Vertrauen in Entscheidungen sind, hilft das Zitieren von konkreten Wahrscheinlichkeiten und Statistiken Menschen nicht, eine Entscheidung nachzuvollziehen. Die wahrscheinlichste Erklärung ist dabei nicht immer die beste Erklärung für einen Menschen. Es müssen Gründe für eine Erklärung angeführt werden.
4. Erklärungen sind sozial. Sie sind Teil eines Wissenstransfers, der in einer Kommunikationssituation stattfindet. Der Erklärende macht seine Erklärung von seinem Gegenüber abhängig.

Erklärungen sind also sehr kontextabhängig und sollten, damit sie funktionieren, in eine Interaktion eingebettet werden. Miller (2018) schlussfolgert mit Blick auf XAI, dass in der Natur von Erklärungen drei Dimensionen zu unterscheiden sind:

1. Fidelity (Grad der Vollständigkeit und Verlässlichkeit der ausgegebenen Erklärung auf Basis des KI-Systems)
2. System Interaction (Erklärungen können interaktiv oder statisch sein. Interaktive Erklärungen können differenzierter auf die Bedürfnisse und Erwartungen von Usern eingehen)
3. Format (Die Form, in der eine Erklärung dargestellt ist. Statistiken, Text, Visualisierungen, ein Mix aus Darstellungsformen)

Diese drei Dimensionen wurden bereits in wissenschaftlichen Arbeiten zu XAI untersucht und sollen im Folgenden näher ausgeführt werden.

Bezüglich der Vollständigkeit wird in der XAI der Umfang einer Erklärung unterschieden, in globale und lokale (Erklärungen) bzw. in Modell- und Instanzebene. Auf globaler Ebene bezieht sich eine Erklärung auf das gesamte KI-System und versucht eine umfassende Darstellung der Entscheidungslogik und der wichtigsten Merkmale des Systems zu liefern. Diese globalen Erklärungen sind allgemeiner und gelten für alle Instanzen, die vom Modell verarbeitet werden. Sie können beispielsweise Informationen über die Gewichtung der Eingabevariablen, die Beiträge einzelner Merkmale zur Vorhersage oder die grundlegenden Muster und Regeln des Systems enthalten (Kamath & Liu, 2021).

Auf lokaler Ebene bzw. der Instanzebene beziehen sich die Erklärungen auf einzelne Instanzen eines Ergebnisses. Lokale Erklärungen sind spezifischer und kontextbezogen, da sie sich auf die

individuellen Merkmale und Eigenschaften der Instanz konzentrieren. Sie können beispielsweise die Gewichtung der Merkmale für die spezifische Vorhersage oder die Änderungen in der Vorhersage liefern, wenn bestimmte Merkmale variiert werden (Kamath & Liu, 2021). Diese Unterscheidung kommt zwar aus einem technischen XAI-Kontext, lässt sich aber auch auf Erklärungen für den Endanwender übertragen. Die Unterscheidung von Erklärungsebenen ermöglicht es, verschiedene Aspekte einer Entscheidungsfindung eines KI-Systems zu erfassen und sowohl eine umfassende Übersicht als auch detaillierte Einblicke in die Funktionsweise der Systeme zu erlangen.

Zu Millers dritter Dimension von Erklärungen liefern Chromik & Butz (2021) eine Übersicht von sieben Interaktionsformen mit XAI. Die einfachste Form ist die Interaktion als (Informations-)Übertragung. Es handelt sich bei dieser Interaktion um eine unidirektionale Kommunikation, bei der eine statische Erklärung präsentiert wird. Das Interface der XAI wird hauptsächlich als Medium für die Übermittlung dieser Erklärung verwendet. Das Ziel dieser Interaktion besteht darin, den Nutzern eine vollständige Erklärung zu geben.

Zwei andere Formen scheinen für den Einsatz in Behörden aber interessanter: Die Interaktion als Dialog und die Interaktion als Tool Use. Das Konzept der „Interaktion als Dialog“ ermöglicht eine dynamische, bidirektionale Kommunikation zwischen Mensch und KI-System. Hierbei können Nutzer spezifische Fragen stellen, auf die die KI mit maßgeschneiderten, verständlichen Erklärungen reagiert. Dieser Ansatz fördert nicht nur ein tieferes Verständnis der KI-Entscheidungen, sondern ermöglicht es den Nutzern auch, ein detailliertes mentales Modell der Funktionsweise des KI-Systems zu entwickeln (Chromik & Butz, 2021).

Parallel dazu betont das Konzept der „Interaktion als Werkzeuggebrauch“ die Rolle des KI-Systems als erweiterndes Instrument, das über seine unmittelbare Funktionalität hinausgeht. Hier wird das KI-System als ein Werkzeug betrachtet, das die Fähigkeiten des Nutzers erweitert, indem es ihm hilft, komplexe Daten zu analysieren und verborgene Muster zu entdecken. Diese Art der Interaktion betont die aktive Rolle des Nutzers, der das KI-System als Werkzeug nutzt, um seine eigenen analytischen Fähigkeiten zu verbessern und tiefere Einblicke in die Daten zu gewinnen.

Die Kombination dieser beiden Konzepte bietet einen vielschichtigen Ansatz für die Interaktion mit XAI-Systemen. Einerseits ermöglicht der dialogbasierte Ansatz eine anpassungsfähige und benutzerorientierte Kommunikation, die es den Benutzern ermöglicht, spezifische Aspekte der KI-Entscheidungsfindung zu erfragen und zu verstehen. Andererseits erweitert die Interaktion als Tool Use das Verständnis und die Fähigkeiten der Nutzer, indem das KI-System als Medium zur Vertiefung ihrer analytischen Kompetenzen dient.

### **Erklärungsbedürfnisse von Anwendern**

Erklärungen sind immer an Menschen gerichtet. Da Menschen sehr unterschiedliche Bedürfnisse haben können, sollten Erklärungen möglichst individuell und passend zum jeweiligen Nutzer und Kontext sein. Zunächst lässt sich festhalten, dass zwischen verschiedenen Typen von Nutzern

unterschieden werden kann, die unterschiedliche Bedürfnisse hinsichtlich Erklärung haben (Liao & Varshney, 2022). Die folgende Grafik nach Arrieta et al. (2020) veranschaulicht die verschiedenen Typen:

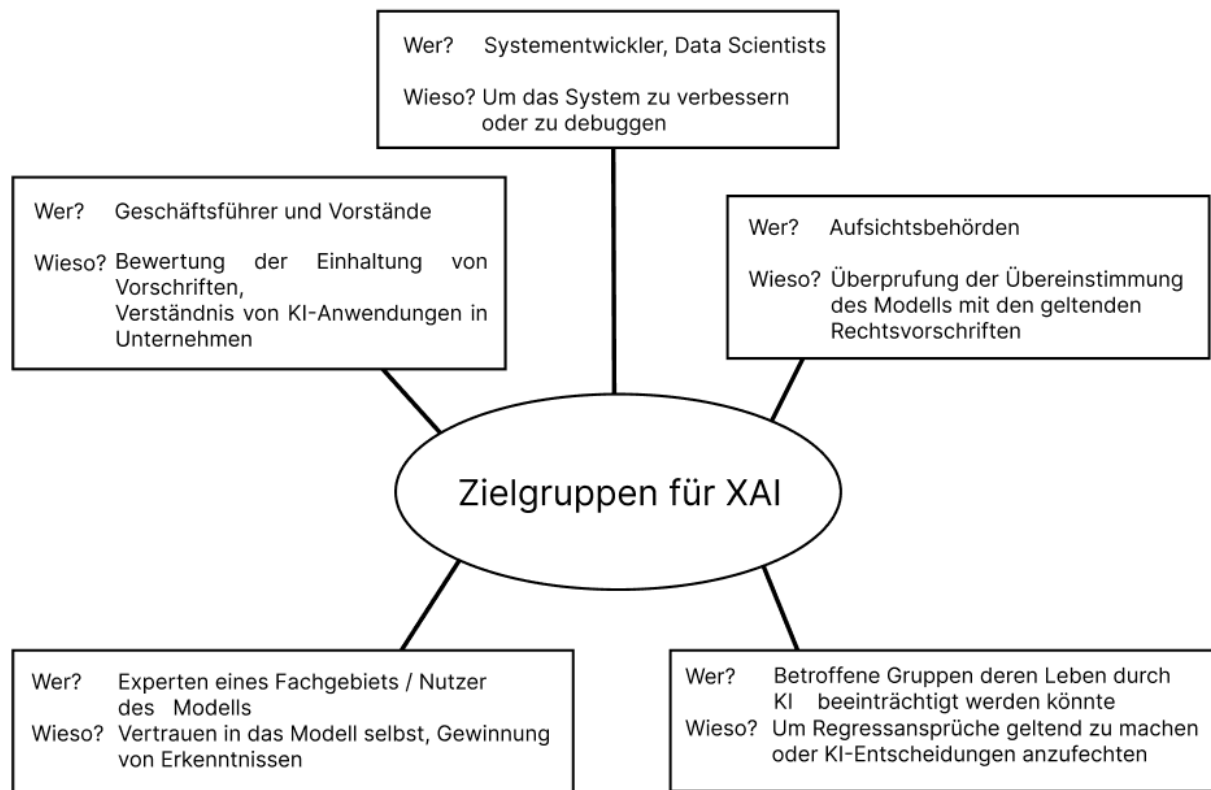


Abbildung 2: Zeigt versch. Ziele für Erklärbarkeit, die von versch. Zielgruppen angestrebt werden. Zwei Ziele stets vorherrschend: Notwendigkeit Modell zu verstehen & Einhaltung v. Vorschriften (Nach einer Grafik aus Arrieta 2020, aus dem Englischen übersetzt)

Da Lebensmittelkontrolleure im Kontext dieser Arbeit Experten auf ihrem Gebiet sind, ist für diese Arbeit vor allem der Nutzertyp *Domain Expert* interessant. Der Nutzertyp „Lebensmittelkontrolleur“ wird in Kapitel 3.4.2 ausführlich beschrieben. Eine Unterscheidung nach Nutzertypen beschreibt jedoch noch nicht den Erklärungsbedarf eines Nutzers. Zwei zentrale Aspekte ermöglichen es, sich den Erklärungsbedürfnissen eines Nutzers anzunähern (Suresh et al., 2021 nach: Liao & Varshney, 2022).

Zum einen ist das Ziel relevant, das ein Nutzer mit der Verwendung eines Systems und der Frage nach einer Erklärung verfolgt. Das Ziel variiert sowohl auf der Ebene der verschiedenen Nutzertypen, wie in Abbildung 2 dargestellt, als auch auf der Ebene des Individuums, das einer Gruppe angehört. Neben den Zielen der Nutzer beim Einsatz von Systemen, stellt sich auch die Frage nach den Zielen, die Systeme oder ihre Entwickler mit Erklärungen für den Nutzer verfolgen.

Laato et al. (2022) sagen, dass die Kommunikation von KI-Systemen für Anwender fünf übergeordnete Ziele verfolgt: (1) Verständlichkeit, (2) Vertrauenswürdigkeit, (3) Transparenz, (4) Kontrollierbarkeit und (5) Fairness. Auf die Punkte Verständlichkeit und Transparenz wird in der Analyse (Kapitel 3.2) noch näher eingegangen. Wie in der Analyse des Nutzungskontextes (Kapitel 3.4) noch genauer gezeigt wird, ist für diese Arbeit das relevante Ziel die Verständlichkeit für Fachexperten (Domain Experts).

Zum anderen ist das Vorwissen bzw. mentale Modell des Nutzers bezüglich des Systems relevant. Unter dem Begriff des *Mentalen Modells* wird in der Mensch-Technik-Interaktion das subjektive Systemverständnis der Nutzer beschrieben (Herczeg, 2009). Das mentale Modell ist in diesem Zusammenhang die dynamische Repräsentation des Systemverständnisses der Nutzer und kann sich durch Erfahrungen und äußere Einflüsse wie z.B. Training verändern (Herczeg, 2009). Bezogen auf Erklärungen von KI-Ergebnissen könnte das bedeuten, dass sich die Vorstellung der Nutzer, wie ein System arbeitet, erst mit wiederholter Darstellung von Erklärungen manifestiert. Deswegen ist es bei der Nutzerkontextanalyse wichtig, das mentale Modell von Nutzern oder Nutzergruppen im Vorhinein zu analysieren, damit das KI-System passende Erklärungen liefern kann.

## 2.5 Fazit

Um sich dem Forschungsthema der Arbeit zu nähern, musste zunächst geprüft werden, unter welchen Bedingungen KI-Systeme in der öffentlichen Verwaltung in Deutschland überhaupt Anwendung finden können. Es hat sich gezeigt, dass diese nur streng reguliert und überwacht eingesetzt werden. Der Mensch muss immer die Kontrolle haben. Deswegen ist es sehr wichtig, dass KI-Systeme in der öffentlichen Verwaltung nachvollziehbar sind. Im zweiten Teil des Kapitels wird der XAI-Ansatz vorgestellt, da Nachvollziehbarkeit dessen Kernelement ist. XAI zeichnet sich dadurch aus, dass er beschreibt, wie Erklärungen von KI-Systemen menschenzentriert sein können.

## 3 Analyse

In diesem Kapitel soll festgestellt werden, welche Anforderungen das zu entwickelnde System erfüllen muss. Dafür wird in Kapitel 3.4 eine Analyse des konkreten Nutzungskontextes der behördlichen Lebensmittelkontrolle und der darin tätigen Nutzer durchgeführt. Die Basis für die Nutzungskontextanalyse bilden u.a. selbstgeführte Interviews, die in Kapitel 3.3 vorgestellt werden.

Weiterhin braucht es für die effektive Umsetzung des Prototyps praktische bzw. anwendungsnahe Lösungen, die in Kapitel 3.2 vorgestellt werden. Eine Literaturrecherche in Kapitel 3.1 zeigt, wie nach den Gestaltungsempfehlungen recherchiert wurde.

### 3.1 Literaturrecherche

Konkrete Guidelines als Zusammenfassungen von Best Practices können für die praktische und zielgerichtete Umsetzung besonders hilfreich sein. Um Faktoren zu ermitteln, von denen zu erwarten ist, dass sie einen Einfluss auf entscheidende Eigenschaften des Systems haben werden, wurde eine Literaturrecherche für wissenschaftliche Arbeiten durchgeführt. Der Fokus liegt auf Faktoren für das Erklärbarmachen von KI-Entscheidungen und ihrer Nachvollziehbarkeit für die Anwender.

Für die Recherche wurde die Datenbank Scopus verwendet, die sich selbst als die größte Abstract- und Zitationsdatenbank von peer-reviewter Forschungsliteratur bezeichnet (What is Scopus about?, 2023). Scopus umfasst bedeutende Veröffentlichungsquellen wie IEEE, SpringerLink und ACM. Da in dieser Datenbank verschiedene Literaturquellen indexiert sind, ermöglicht sie eine schnelle und breit gefächerte Suche und damit eine effektive Erfassung der relevanten Forschungsliteratur.

Die Suchbegriffe sollten dabei im Titel, Abstract oder in den Schlüsselwörtern vorkommen. Mit folgender Sucheingabe wurde nach Guidelines recherchiert:

```
TITLE-ABS-KEY("xai" OR "Explainable AI" OR "transparent AI" OR "interpretable AI" OR "explanation facility" OR "explanation interface" OR "AI explainability" OR "AI transparency" OR "AI interpretability" AND "user interaction" OR "user interface" OR "usability" OR "interactive" AND "guideline" OR "design factor" OR "design recommendation") AND PUBYEAR > 2018
```

Als Ergebnis wurden 131 Vorschläge ausgegeben. Zuerst wurden die Ergebnisse durch Aussortieren unpassender Titel gefiltert. Für die übrig gebliebenen Vorschläge wurden die Abstracts überflogen. Am Ende blieb 1 Ergebnis übrig.

Da nur eine passende Arbeit ausfindig gemacht wurde, die eine Guideline enthielt, wurde für das Aufstellen konkreter Anforderungen weitere Literatur hinzugezogen.

## 3.2 Faktoren für die Gestaltung von XAI-Systemen

In Kapitel 2.4 wurden verschiedene Aspekte von Erklärungen im Kontext von XAI vorgestellt. Sie zeigen, dass die Vermittlung von Erklärungen an Nutzer eine zentrale Herausforderung für die Gestaltung von menschenzentrierten KI-Systemen darstellt. Die im XAI-Kontext relevanten theoretischen Aspekte werden im folgenden Abschnitt um konkrete Gestaltungsempfehlungen für XAI-Systeme ergänzt.

### **Bedeutung von Kompromissen für Gestaltung von XAI-Systemen**

Laato et al. (2022) haben eine systematische Literaturanalyse durchgeführt, um zu untersuchen, wie KI-Systeme und ihre Entscheidungen für Nutzer erklärt werden sollten. Die Autoren haben mehrere Gestaltungsempfehlungen zusammengetragen und eine Guideline formuliert. Da Nutzer eine heterogene Gruppe bilden, gibt es keine universelle Lösung, die allen Anforderungen gerecht wird. Daher ist es nötig, Kompromisse bei der Gestaltung von Erklärungen von KI-Systemen einzugehen. Kompromisse (englisch: Trade-Offs) sind zentral für die Gestaltung von erklärenden Systemen. Ein Beispiel für solch einen Kompromiss ist die Balance zwischen Vollständigkeit und Verständlichkeit: Je umfassender ein System erklärt wird, desto komplexer wird die Erklärung für den Nutzer. Soll eine Erklärung eines Systems aber nachvollziehbar sein, ist es evtl. notwendig, nicht jedes Detail zu erklären. Das Weglassen von Details für eine Optimierung auf Verständlichkeit kann aber dazu führen, dass das Vertrauen der Nutzer in die Erklärung beeinträchtigt wird.

### **Faktoren für Verständlichkeit und Transparenz**

In Kapitel 2.4 wurden fünf Ziele genannt, die mit erklärender Kommunikation von Systemen (bzw. den Entwicklern) verfolgt werden können. Die Ziele Verständlichkeit und Transparenz sind für diese Arbeit besonders interessant. Die dazugehörigen Faktoren werden in diesem Abschnitt betrachtet. Die zwei Ziele stehen in einem engen Zusammenhang und beeinflussen sich gegenseitig, da die wahrgenommene Transparenz mit der Fähigkeit der Nutzer zusammenhängt, den Inhalt zu erfassen.

Verständlichkeit meint das Ausmaß, in dem die Nutzer in der Lage sind, sich ein genaues mentales Modell von der Funktionsweise des KI-Systems zu machen. Die Literatur zeigt, dass das technische Wissen, die früheren Vorstellungen und die mentalen Fähigkeiten der Nutzer berücksichtigt werden müssen, wenn ihnen KI-Systeme erklärt werden. Je besser die Erklärungen auf die Vorstellungen und mentalen Modelle der Nutzer abgestimmt waren, desto besser verstanden sie die Erklärungen (Laato et al., 2022). Dies war besonders relevant für KI-Systeme, bei denen Nutzer mit wenig technischem Fachwissen eine besonders hohe Wahrscheinlichkeit haben, falsche Vorstellungen darüber zu haben, wie KI-Systeme zu Schlussfolgerungen kommen. Daher müssen sich XAI-Systeme darauf konzentrieren, Erklärungen zu liefern, die intuitiv sind, d. h. Visualisierungen (non-verbale Kommunikation) und sogar Metaphern zu verwenden (Laato et al., 2022).

Auch die verbale Kommunikation muss an die Bedürfnisse der Nutzer angepasst werden. So behindern eine zu vereinfachte und zu komplizierte Sprache die Fähigkeit der Nutzer, ein System nachzuvollziehen. Zu komplexe und detaillierte Erklärungen könnten bei den Teilnehmern eine Informationsüberlastung hervorrufen. Außerdem ist die kontextuelle Genauigkeit wichtiger als die Länge der Erklärung (Laato et al., 2022). Zusammenfassend lässt sich sagen, dass die Verständlichkeit von KI-Systemen vom System selbst, von der verbalen und nonverbalen Kommunikation sowie von den Erfahrungen und dem Vorwissen der Nutzer abhängt.

Transparenz kann als Bereitstellung von mehr Informationen über das System für die Nutzer definiert werden. Für die Transparenz von KI-Systemen und ihren verwendeten Daten gilt, dass eine angemessene Transparenz angestrebt werden sollte. Eine vollständig transparente und umfassende Erklärung ist jedoch nicht immer die hilfreichste Lösung für die Nutzer. In der Debatte über Transparenz geht es nicht nur darum, was erklärt werden kann, sondern auch darum, welche Systeminformationen für die Nutzer tatsächlich relevant sind (Eiband 2018, zitiert nach Laato et al., 2022).

Insgesamt haben die Autoren 16 Gestaltungsempfehlungen ermittelt, die in vier Kategorien eingeteilt sind: Neben der Kategorie „allgemeine Empfehlungen“ außerdem Empfehlungen, die sich auf das "Wann", "Was" und "Wie" beziehen, um KI-Systementscheidungen zu erklären. Abbildung 3 stellt diese in ihrer Gesamtheit dar.

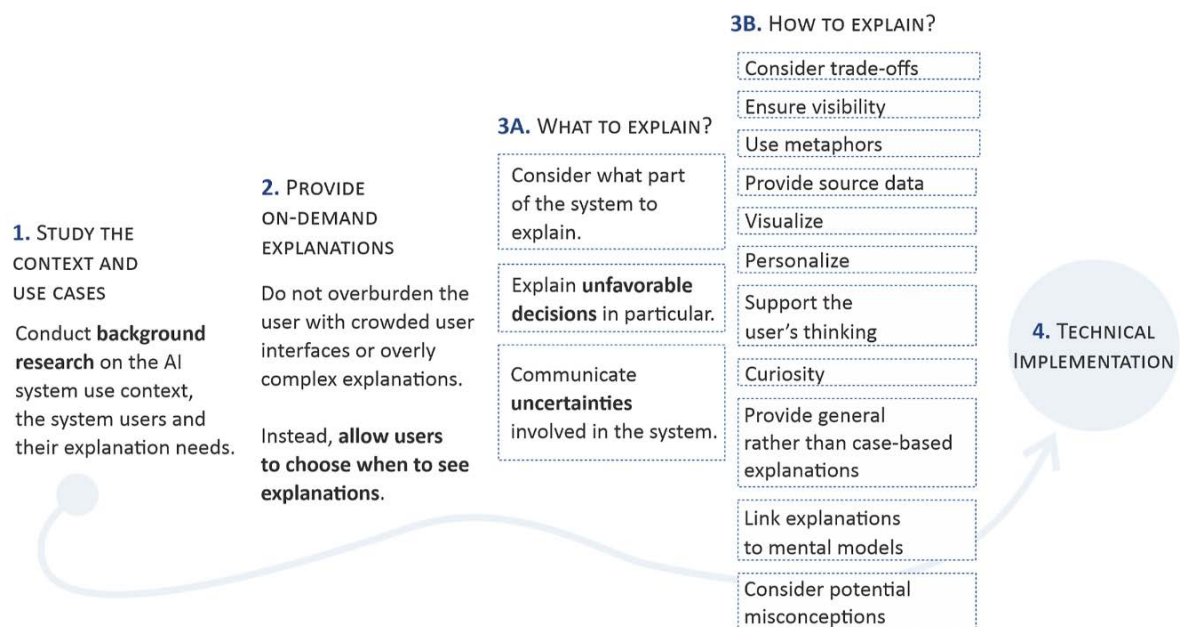


Abbildung 3: Ein Design Framework für die Kommunikation von KI-Systemen, dass zu den 5 Zielen Verständlichkeit, Vertrauen, Transparenz, Beherrschbarkeit und Fairness passt (aus Laato et al., 2022)

### **Faktoren für die Nachvollziehbarkeit**

Weitere Anforderungen für nachvollziehbare KI-Systeme wurden in einer Studie gefunden, die bereits verwendete KI-Systeme in der öffentlichen Verwaltung in Deutschland untersucht haben (Engelmann & Puntschuh, 2020). Die Autoren haben eine Kriterienmatrix mit neun Kriterien für Best Practices beim KI-Einsatz in der Verwaltung erarbeitet. Eines dieser Best Practices betrifft die Nachvollziehbarkeit von KI-Systemen und ihren verwendeten Daten. Es wird gefordert, dass die verwendeten Daten für jede Entscheidung nachvollziehbar dokumentiert werden. Der Einsatz des Systems sollte gekennzeichnet und gegenüber den Nutzern und Betroffenen transparent gemacht werden. Bei Entscheidungen, bei denen das System zum Einsatz kam, sollten die Betroffenen unaufgefordert über dessen Einsatz, die zugrunde liegenden Daten und die relevanten Faktoren aufgeklärt werden. Diese Informationen sollten zielgruppenspezifisch und verständlich aufbereitet sein, um sie für Betroffene, Führungskräfte und externe Stakeholder zugänglich zu machen.

Besonders die Verwendung von Erklärungen, die über reine Erklärtexte hinausgehen und die Nachvollziehbarkeit einzelner Entscheidungen des KI-Systems unterstützen, ist laut den Autoren erstrebenswert. Dies geschieht, indem die jeweils für die Entscheidung relevanten Faktoren hervorgehoben und eine individuelle Wahrscheinlichkeit der Korrektheit ausgegeben wird (Engelmann & Puntschuh, 2020). Die genannten Faktoren für Nachvollziehbarkeit sind auch für den Bereich der Lebensmittelkontrolle relevant, da Lebensmittelkontrolleure für mögliche Fehlentscheidungen, die aus einer Fehlinterpretation des KI-Systems resultieren, haftbar gemacht werden könnten.

### **Faktoren für den Umfang und die Vollständigkeit der Erklärungen**

Da in der Recherche für diese Arbeit nur wenige konkrete Gestaltungsempfehlungen für XAI-Interfaces gefunden wurden, wurde zusätzlich Literatur aus dem medizinischen Kontext herangezogen. Dieser Bereich der XAI-Forschung beschäftigt sich mit der nachvollziehbaren Gestaltung von Benutzeroberflächen für Nutzer und ist besonders relevant, da hier Anwender mit überwiegend nicht-technischem Hintergrund KI-Systeme in risikobehafteten Kontexten nutzen müssen.

Es wurde eine Studie gefunden, die u.a. das Ziel hat, die Auswirkungen von Erklärungen auf die subjektive Wahrnehmung von Nachvollziehbarkeit von Nutzern eines Systems zu untersuchen. Dafür haben die Autoren ein Simulationssystem entwickelt, mit der ein Insulinbedarf berechnet werden sollte. Die dem Algorithmus zur Verfügung stehenden Informationen wurden den Studienteilnehmern in unterschiedlichem Umfang offengelegt, um drei verschiedene Versuchsbedingungen zu schaffen. Es wurde untersucht, ob eine größere Menge an Informationen zu einer höheren erlebten Nachvollziehbarkeit führt. Außerdem wurde analysiert, inwieweit das wiederholte Betrachten von erklärenden Informationen zu einer Erhöhung der erlebten Nachvollziehbarkeit führen kann (Schrills & Franke, 2022).



Den Nutzern wurden die Parameter präsentiert, die als Basis für die Berechnung des Insulinbedarfs dienen. Anschließend sollten die Probanden die Berechnung des Systems auf Grundlage der zur Verfügung stehenden Informationen vorhersagen. Die subjektiv wahrgenommene Nachvollziehbarkeit wurde mit der Leistung der zu erledigenden Aufgaben verglichen.

In ihrem Experiment korreliert die eigene Einschätzung der Teilnehmer über die Nachvollziehbarkeit des Systems nicht mit ihrer Fähigkeit, die Berechnung des Systems vorherzusagen. Die Autoren sprechen von einem besorgniserregenden Zusammenhang, bei dem im günstigsten Fall falsche Erwartungen entstehen und das Vertrauen in das System verloren geht. Eine schwerwiegendere Folge könnte eine Fehleinschätzung der Systemleistung in extremen Situationen sein oder die Entwicklung von übersteigertem Selbstvertrauen in die eigenen Fähigkeiten mit dem System (Schrills & Franke, 2022).

Die Untersuchung zeigt, dass die Wirkung verschiedener Grade der Offenlegung von Informationen möglicherweise mehrere Wiederholungen benötigt, bevor sie sich manifestiert. Die Ergebnisse deuten darauf hin, dass Entwickler für eine verantwortungsvolle Gestaltung von XAI-Systemen Vorhersageaufgaben nutzen könnten, um die erfahrene Nachvollziehbarkeit zu kalibrieren (Schrills & Franke, 2022).

Bei der Entwicklung von XAI-Systemen ist es wichtig, sowohl die Effekte von Erklärungen auf die wahrgenommene Nachvollziehbarkeit als auch auf die Arbeitsbelastung und Vorhersageleistung zu beachten. Die bloße Bereitstellung zusätzlicher Informationen kann nicht als geeignete Methode zur Verbesserung des Nutzererlebnisses oder als Grundlage für die Mensch-KI-Interaktion gesehen werden. Die Tatsache, dass dieser Ansatz keinen signifikanten Vorteil für die Teilnehmer aus der Gruppe der Bedingung mit der höchsten Informationsoffenlegung bot, unterstreicht den Bedarf an menschenzentrierter Forschung im XAI-Bereich (Schrills & Franke, 2022).

### **Tabellarische Zusammenfassung der Erkenntnisse**

Die in diesem Abschnitt dargestellten Faktoren werden zur besseren Übersicht hier nochmal tabellarisch aufgelistet:

<b>Zieleigenschaft (Betreffende Eigenschaft)</b>	<b>Beeinflussender Faktor</b>	<b>Beschreibung</b>
<b>Verständlichkeit</b>	personalisierte und bedarfsgerechte Erklärungen	Die angebotenen Erklärungen sollten personalisiert und bedarfsgerecht sein.

	Erklärbarkeit von Schlüsselfunktionen	Das Erklären von Schlüsselfunktionen ist besser für die Verständlichkeit als das Erklären des gesamten Systems.
	Eingehen von Kompromissen	Es ist notwendig Kompromisse einzugehen bei der Gestaltung von Erklärungen für KI-Systeme.
	Mentale Modell der Nutzer beachten	Das technische Wissen, die früheren Vorstellungen und die mentalen Fähigkeiten der Endnutzer haben Einfluss auf das Verständnis für ein System.
	Intuitive Erklärungen (non-verbale Kommunikation)	Erklärungen sollten intuitiv sein, d.h. auch Visualisierungen und sogar Metaphern nutzen.
	Verbale Kommunikation	Die verbale Kommunikation muss an die Bedürfnisse der Nutzer angepasst werden, d.h. nicht zu einfach und nicht zu kompliziert.
	kontextuelle Genauigkeit	Die kontextuelle Genauigkeit ist wichtiger als die Länge einer Erklärung.
<b>Nachvollziehbarkeit</b>	Angemessenheit der Transparenz	Es sollte eine angemessene Transparenz angestrebt, da eine völlig transparente, umfassende Erklärung ist nicht immer die nützlichste Lösung ist.
	Transparenz der verwendeten Daten	Die verwendeten Daten sollten für jede Entscheidung nachvollziehbar dokumentiert werden.
	Verfügbarkeit von Informationen zur Funktionsweise	Informationen, die die Gesamtfunktionsweise des Systems beschreiben, sollten (öffentlich) verfügbar sein, um Nachvollziehbarkeit zu

		gewährleisten.
	Transparenz der Verwendung von KI	Der Einsatz eines Systems sollte gekennzeichnet und gegenüber den Nutzern (und Betroffenen) transparent gemacht werden.
	Transparenz der verwendeten Daten	Die Nutzer (und Betroffenen) sollten ohne Aufforderung über den Einsatz, die zugrunde liegenden Daten und die relevanten Faktoren aufgeklärt werden.
	Hervorheben relevanter Faktoren	Die Nachvollziehbarkeit von einzelnen Entscheidungen des KI-Systems kann erhöht werden, indem die jeweils für die Entscheidung relevanten Faktoren hervorgehoben werden.
<b>Informations- umfang</b>	Empfundene Nachvollziehbarkeit	Wenn Nutzer ein System als nachvollziehbar empfinden, heißt es nicht, dass sie dann das System verstanden haben.
	Wiederholung von Erklärungen	Erklärungen werden nicht immer sofort verstanden und sollten wiederholt werden können.
	Ein Mehr an Informationen	Die alleinige Weitergabe von Zusatzinformationen bedeutet nicht, dass der Nutzer das System besser versteht.

*Tabelle 1: Ermittelte Faktoren für die Nachvollziehbarkeit eines Systems*

### 3.3 Datenquellen der Nutzungskontextanalyse

Nachdem im vorherigen Kapitel allgemeine Faktoren für die Verständlichkeit eines Systems gesammelt und erklärt wurden, liegt im folgenden Schritt der Fokus darauf, den Nutzungskontext der Lebensmittelkontrolle zu verstehen. Es wurden dazu verschiedene Quellen herangezogen, die Informationen zum Arbeitsfeld und der Zielgruppe bieten.

Zu Beginn der Recherche wurden die Allgemeine Verfahrensverordnung Rahmen-Überwachung (AVV Rahmen-Überwachung – AVV RÜb, 2021) und die dazugehörigen Anlagen analysiert. Die Verordnungen gaben einen theoretischen Einblick in die Aufgaben, den Umfang und die Prozesse der Arbeit von Lebensmittelkontrolleuren. Diese bilden den rechtlichen Rahmen für die Bewertung und Kategorisierung von Betrieben, wobei die praktische Arbeit der Kontrolleure teilweise von diesen Vorgaben abweichen kann.

Für praktische Einblicke wurden Ergebnisprotokolle von zwei Interviews mit Lebensmittelkontrolleuren, die im Rahmen des Forschungsprojekts KI-DemÖ der Universität Lübeck durchgeführt wurden, herangezogen. Die Zusammenfassungen dieser Interviews sowie ein Gespräch mit dem wissenschaftlichen Mitarbeiter, der die Interviews geführt hat, boten wertvolle Informationen. Insbesondere, da zunächst unklar war, ob ein eigenständig für diese Arbeit geführtes Interview mit einem Lebensmittelkontrolleur möglich sein würde. Die zwei Interviews wurden jedoch lediglich per Telefon geführt und konnten daher keine Einblicke in die konkrete Planung von Kontrollen oder der Arbeit mit der in der Behörde verwendeten Software bieten.

Um diese Lücke zu füllen und ein tieferes Verständnis für die Nutzer zu bekommen, wurde zusätzlich ein eigenes Interview mit einem Lebensmittelkontrolleur organisiert und durchgeführt. Dieses fand vor Ort in den Büros der Kieler Lebensmittelüberwachung statt. Das Interview lieferte einen direkten Einblick in die derzeit von der Behörde verwendeten Software und wie damit die Planung der Kontrollen umgesetzt wird. Darüber hinaus konnten weitere interessante und wichtige Informationen zum Nutzungskontext gesammelt werden. Die Ergebnisse der Analyse des Nutzungskontextes werden im folgenden Kapitel (3.4) vorgestellt.

### 3.4 Nutzungskontextanalyse

Nachdem im ersten Teil des Analysekapitels mittels Literaturrecherche (Kapitel 3.1) Forschungsarbeiten und Studien untersucht wurden, um wichtige Faktoren (Kapitel 3.2) für das zu entwickelnde System zu ermitteln, wird im folgenden Teil die Nutzungskontextanalyse beschrieben. Die Analyse der verschiedenen Facetten des Nutzungskontextes ist ein zentraler Bestandteil des nutzerzentrierten Entwicklungsprozesses, mit dem Ziel, Eigenschaften und Anforderungen zu ermitteln, um das System optimal an die Fähigkeiten und das mentale Modell der Kontrolleure anzupassen.

### 3.4.1 Die Aufgaben der Lebensmittelkontrolleure

Im folgenden Abschnitt wird die durchgeführte Aufgabenanalyse beschrieben. Die Hauptaufgaben von Lebensmittelkontrolleuren sind, Verbraucher vor Gesundheitsgefahren, Irreführung und Täuschung zu schützen und ihnen zielgerichtete Informationen bereitzustellen. Da Lebensmittelunternehmer – also jene Personen oder Unternehmen, die Lebensmittel herstellen, verarbeiten und in Umlauf bringen – für die Lebensmittelsicherheit verantwortlich sind, führt die amtliche Lebensmittelüberwachung eine Art „Kontrolle der Kontrolle“ durch. Sie überprüft stichprobenartig und risikoorientiert, ob die Lebensmittelbetriebe die gesetzlichen Vorschriften einhalten und sicherstellen, dass die von ihnen in den Verkehr gebrachten Erzeugnisse keine Gefahr für die Gesundheit der Verbraucher darstellen. Eine weitere Aufgabe der Lebensmittelüberwachung ist der Vollzug von geeigneten Maßnahmen bei Verstößen der Betriebe gegen das Lebensmittelrecht.

Dabei ist der Bereich, in dem kontrolliert wird, sehr vielfältig. Neben den typischen Betrieben, wie Lebensmittelproduktionsbetrieben, Gaststätten, Lebensmitteleinzelhandel oder Großhandel, werden auch Bedarfsgegenstände mit Lebensmittelkontakt oder Körperkontakt wie Tabakerzeugnisse, Kosmetik, Schmuck, Spielzeug, Geschirr oder Tattoofarben kontrolliert. Da Lebens- und Futtermittel bis zu ihrem Ursprung rückverfolgbar sein müssen, gilt der Rechtsrahmen für die gesamte Erzeugungskette. Die meiste Arbeitszeit (etwa 70%) wird auf die oben beschriebenen Aufgaben im Außendienst verwendet, die restliche Zeit verbringen Lebensmittelkontrolleure im Büro mit Aufgaben wie der Planung ihrer Einsätze.

#### **Der theoretische Prozess der Bewertung von Betrieben**

Die Beantwortung der Frage, in welchen Abständen und wie Betriebe kontrolliert werden, ist durch die AVV RÜb geregelt. Sie setzt geltendes EU-Recht um, da innerhalb der EU das Lebensmittelrecht weitgehend harmonisiert ist (AVV Rahmen-Überwachung – AVV RÜb, 2021). Die Beurteilung, in welchem Abstand eine Betriebskontrolle stattfindet, erfolgt risikoorientiert. Die AVV RÜb beinhaltet dafür unter anderem einen bundesweit geltenden Kriterienkatalog zur Beurteilung des Risikos eines Betriebs. Konkret geht es darum, Betriebe mit höherem Risiko zu identifizieren und diese entsprechend intensiver zu beobachten und häufiger zu kontrollieren als Betriebe mit geringerem Risiko. Für jedes Kriterium sind Punktzahlen definiert, die bei entsprechenden Ausprägungen des Kriteriums zu vergeben sind. Aus der Gesamtpunktzahl eines Betriebs ergibt sich, wie häufig dieser kontrolliert werden muss. Neben der Art des Betriebs und der Produkte, fließen weitere Bewertungsmerkmale, wie das bisherige Abschneiden des Unternehmens bei Kontrollen, die Verlässlichkeit von Eigenkontrollen und das Hygienemanagement ein. Für die Umsetzung einer Beurteilung ist in Anlage 1 der AVV RÜb ein Beurteilungsbogen spezifiziert. Nach diesem werden zunächst Daten, wie Betriebsname, Datum der Beurteilung des Betriebs und Name der beurteilenden Person, festgehalten.

Vor einer betriebspezifischen Beurteilung des Betriebs vor Ort, wird zunächst eine Ersteinstufung gemacht. Durch die Ersteinstufung wird eine erste Kontrollhäufigkeit für den

betreffenden Betrieb unter der Annahme einer zufriedenstellenden Einhaltung der lebensmittelrechtlichen Anforderungen ermittelt. Beurteilungsmerkmale, wie Betriebs- und Produktrisiko, sind statisch und ändern sich nur, wenn sich der Betriebs- oder Produkttyp ändert. Dynamische Merkmale, wie bisheriges Verhalten des Unternehmens bekommen bei der Ersteinstufung eine mittlere Beurteilung. Eine Feineinstufung erfolgt bei der ersten Beurteilung im Betrieb. In der Feineinstufung werden die dynamischen Merkmale nun anhand von Beurteilungskriterien eingestuft (AVV Rahmen-Überwachung – AVV RÜb, 2021).

Es gelten festgelegte Risikoklassen, Kontrollhäufigkeiten sowie erreichbare Punkte. Die Gesamtpunktzahl, die ein Betrieb erreichen kann, liegt zwischen 0 und 200 Punkten. Die Bewertung mit der besten Beurteilungsstufe (0 Punkte) erfolgt dann, wenn die einschlägigen Rechtsgrundlagen vollumfänglich erfüllt sind und der Betrieb in einer entsprechenden Kategorie einzuordnen ist. Je nach Punktzahl wird der Betrieb in eine Risikoklasse eingestuft und anhand dieser wird die Kontrollfrequenz der Regelkontrollen bestimmt, die sich zwischen mindestens wöchentlich bis fünfjährlich bewegen kann (AVV Rahmen-Überwachung – AVV RÜb, 2021).

Außerhalb der festen Kontrollzyklen sind zusätzlich stattfindende anlassbezogene Kontrollen möglich, z.B. aufgrund von Nachkontrollen, bei Verdachtsfällen oder Beschwerden aus der Bevölkerung und bei betrieblichen Änderungen.

Der Zeitwand für eine Kontrolle einschließlich Vor- und Nachbereitung, unterscheidet sich je nach Größe und Art des Betriebs. Außerdem werden bei manchen Kontrollen nur Teile eines Betriebs oder eine Auswahl der Kontrollkriterien geprüft (Lebensmittelkontrolleur, persönliche Kommunikation, 14.07.23).

### **Die praktische Planung von Kontrollen**

Lebensmittelkontrolleure nutzen in der Praxis die Softwarelösung Balvi IP, die Teile des Bewertungs- und Kontrollplanungsprozesses automatisiert, einschließlich der Berücksichtigung von Fristen und Risikoeinschätzungen bei der Terminplanung. In Balvi IP können Nutzer den gewünschten Zeitraum festlegen, für den die zu kontrollierenden Betriebe angezeigt werden. Dies kann ein festgelegter Zeitraum wie der aktuelle Monat oder ein individuell gewählter Zeitraum sein. Der interviewte Kontrolleur wählt beispielsweise den Zeitraum seit dem letzten obligatorischen Bezirkswechsel bis einen Monat in die Zukunft, um mögliche Versäumnisse des Vorgängers zu überprüfen, und lässt sich dann die Liste der Betriebe vom 1. Januar 2019 bis zum Ende des aktuellen Monats anzeigen. Balvi IP stellt daraufhin eine Liste bereit, die nach verschiedenen Kriterien, beispielsweise alphabetisch, sortiert werden kann. Wer noch kein Tablet nutzt, druckt sich diese Liste aus. Es ist wichtig zu beachten, dass die Software nicht darauf aufmerksam macht, wenn bei der Abfrage ein Zeitraum übersprungen wird, in dem ein Betrieb eigentlich hätte kontrolliert werden müssen.

### **Die Rolle von Daten und deren Einfluss auf Entscheidungen in der Lebensmittelkontrolle**

Wie die Anforderungen für ein XAI-System gezeigt haben, ist das transparente Zurverfügungstellen der Daten, die für eine KI-Berechnung verwendet werden, eine wichtige

Eigenschaft von XAI-Systemen. Daher ist die Frage, wie in den aktuellen Planungs- und Kontrollprozessen mit vorhandenen Daten, wie Kontrollberichten, umgegangen wird. Auf die Frage, ob Kontrollprotokolle oder andere Daten herangeholt werden bei der Planung bzw. vor einer Kontrolle, antwortete der interviewte Lebensmittelkontrolleur, dass es möglich wäre, er das aber nicht tun würde, da Kiel relativ übersichtlich sei. Entscheidend sei die „Optik vor Ort“. Die Kontrolle vor Ort entscheidet, ob ein Betrieb eine neue Kontrollfrequenz bekommt oder darüber hinaus weitere Maßnahmen ergriffen werden. Das lässt darauf schließen, dass vergangene Kontrollen bei der Beurteilung des Risikos entweder keine große Rolle spielen bzw. hauptsächlich der Status Quo über zu treffende Bewertungen und Maßnahmen entscheidet. Es stellt sich die Frage, ob sich Kontrolleure an vergangene Mängel eines Betriebs erinnern können und keine zusätzliche Einsicht in hinterlegte Daten benötigen.

Eine interessante Beobachtung, die sich im Interview gezeigt hat, ist eine Funktion der Software Balvi IP, die es erlaubt, bei auffälligen Betrieben einen Haken zu setzen. Dabei liegt es im Ermessen des Kontrolleurs, ob dieser Haken gesetzt wird. Es ist ebenfalls die Entscheidung des Kontrolleurs, ob eine Regelverletzung als Verstoß angesehen und dementsprechend im System vermerkt wird. Falls Mängel direkt vor Ort behoben werden können, besteht die Möglichkeit, dass diese nicht im System erfasst werden, da sie nicht notiert wurden.

### **Vollständigkeit analoger und digitaler Daten**

Der Umfang bzw. die Vollständigkeit der digitalen Daten, wie etwa die aus Kontrollprotokollen, hängt davon ab, ob die Kontrolleure ihre Daten in das Softwaresystem übertragen. Dabei ist die Digitalisierung aller Daten eigentlich Vorgabe im Amt. Im Gespräch mit dem Lebensmittelkontrolleur wurde deutlich, dass die Disziplin diesbezüglich nicht bei allen Kollegen gleich hoch ist. Dies ist auch der Fall, wenn die An- bzw. Abmeldedaten der Gewerbebehörde eingepflegt werden müssen. Jeder Kontrolleur muss die in seinem Bezirk liegenden Betriebe selbst heraussuchen und einpflegen. Bei der Übertragung kann es zu Ungenauigkeiten in den Daten kommen. Es wurde jedoch angemerkt, dass eine Verbesserung der Datenqualität und -menge zu erwarten ist, sobald alle Kollegen auf Tablets umgestiegen sind, die es ihnen ermöglichen, Notizen direkt digital zu erfassen und mit dem Softwaresystem zu synchronisieren. Auch soll im Herbst 2023 für die Kieler Behörde die elektronische Akte eingeführt werden.

### **Das derzeitige Angebot von Erklärungen im verwendeten Softwaresystem**

Die Software Balvi IP bietet aktuell verschiedene Arten von Erklärungen und Hilfestellungen, um den Nutzern den Umgang mit dem System zu erleichtern. Eine der zentralen Ressourcen ist ein digitales Handbuch, das umfassende Informationen über die Funktionsweise der Software enthält. Dieses Handbuch dient als eine Art Nachschlagewerk.

Zusätzlich verfügt Balvi IP über eine spezielle Schulungsebene im System. Diese Ebene ist so konzipiert, dass Nutzer alle Funktionen der Software ausprobieren können, ohne dabei Risiken einzugehen, etwas Fehlzusteuern oder bestehende Daten zu beeinträchtigen. Durch diese Trainingsumgebung haben die Lebensmittelkontrolleure die Möglichkeit, sich mit den

verschiedenen Features und Abläufen vertraut zu machen, was eine praktische Lernerfahrung bietet.

### **Bisher verwendete Datenquellen**

Um für die spätere Gestaltung des XAI-Systems zu wissen, welche Art von Daten integriert und repräsentiert werden müssen, wurde dem Kontrolleur eine Frage nach möglichen nutzbaren Datenquellen für die Lebensmittelkontrolle gestellt. Die Vielfalt der bisher verwendeten Datenquellen ist für den normalen Kontrolleur dabei übersichtlich. Eine wesentliche Informationsquelle sind die Kontrollprotokolle, die detaillierte Aufzeichnungen über frühere Überprüfungen und Befunde enthalten. Eine weitere wichtige Datenquelle stellen die Daten aus der Gewerbeanmeldung dar. Hier können Kontrolleure auf Kontaktdaten der Unternehmen zugreifen und erfahren, welche Geschäftsaktivitäten die Unternehmer planen. Diese Informationen sind besonders nützlich, um ein vollständiges Bild der zu kontrollierenden Betriebe zu erhalten.

Darüber hinaus nutzen Kontrolleure das Internet als Ressource, um ergänzende Informationen wie Kontaktdaten oder Öffnungszeiten zu recherchieren, falls diese in den offiziellen Dokumenten fehlen sollten. Laut Aussage eines Lebensmittelkontrolleurs sind diese Quellen in der Regel ausreichend für die Arbeit. Zusätzliche Daten werden selten benötigt, da sich der Großteil der erforderlichen Informationen direkt aus den durchgeführten Kontrollen ergibt.

### **3.4.2 Die Nutzer**

Der folgende Abschnitt beschreibt die Nutzer als Zielgruppe des zu entwickelnden Systems, die ausschließlich aus Lebensmittelkontrolleuren besteht. In Deutschland gibt es schätzungsweise 2.500 Lebensmittelkontrolleure in über 400 Lebensmittelüberwachungsbehörden, die zusammen mit Tierärzten und Lebensmittelchemikern etwa 1,2 Millionen registrierte Lebensmittelunternehmen überwachen (BVLK, 2022). Typischerweise sind Kontrolleure alleinverantwortlich für ein eigenes Kontrollgebiet zuständig. Ein Kontrollgebiet kann in Schleswig-Holstein 350-400 Betriebe umfassen. Damit arbeiten sie hauptsächlich alleinverantwortlich (Lebensmittelkontrolleur, persönliche Kommunikation, 14.07.23).

Die Gruppe der Lebensmittelkontrolleure ist keine homogene Gruppe. Ihre Altersstruktur entspricht der typischen Altersspanne von Berufstätigen, also ungefähr zwischen 20 und 67 Jahren. Trotz des gleichen Berufsbildes variieren ihre beruflichen Hintergründe. Zwar ist ihnen fast allen gemein, dass sie den Abschluss als Lebensmittelkontrolleur im Rahmen einer Weiterbildung nach dem Berufsbildungsgesetz erhalten haben und weitere Voraussetzung meist eine Ausbildung in einem relevanten Beruf ist, wie z.B. ein Meister im Lebensmittelhandwerk. Einige steigen auch über die Beamtenlaufbahn in den Beruf ein.

In der Weiterbildung werden Lebensmittelkontrolleure hinsichtlich des Umgangs mit EDV-Systemen geschult. Außerdem bekommen einzelne Kollegen Schulungen zu den genutzten



Softwaresystemen, z.B. bei Veränderungen, und können ihr Wissen als Multiplikator an andere Kollegen weitergeben.

Aufgrund der verschiedenen Demografien und beruflichen Hintergründe ist anzunehmen, dass das Vorwissen und die Erfahrungen im Umgang mit Computersystemen stark variieren.

Auf die Frage, ob er neben seiner Tätigkeit als Lebensmittelkontrolleur noch eine andere Funktion in seinem Team ausübt, antwortete der interviewte Lebensmittelkontrolleur, dass er ebenfalls als Admin tätig ist. In dieser Funktion agiert er als Bindeglied bei IT-bezogenen Anliegen. Da sein Ruhestand kurz bevorsteht, wird aktuell nach einem Nachfolger gesucht. Trotz des Vorhandenseins junger Masterabsolventen unter seinen Kollegen, zeigt niemand Interesse, diese Rolle zu übernehmen. Dies könnte Rückschlüsse auf die Einstellung des Kollegiums auf IT-bezogene Themen zulassen. Es könnte darauf hinweisen, dass die einhergehende Verantwortung als unattraktiv empfunden wird oder die spezifischen Fähigkeiten, die für diese Position erforderlich sind, nicht vorhanden sind. Auch könnte eine Rolle spielen, wenn die zusätzliche Tätigkeit des IT-Admins nicht vergütet wird oder die Arbeitsbelastung mit der zusätzlichen Tätigkeit noch weiter steigt.

### 3.5 Fazit der Analyse

Die Analyse hatte das Ziel, die Anforderungen an das entstehende System zu ermitteln. Dabei wurden Anforderungen für XAI-Systeme aus der Forschungsliteratur gesammelt und Interviews geführt, die Einblicke in den Nutzungskontext der Lebensmittelüberwachung boten. Es wurde der rechtliche Rahmen auf den vorgeschriebenen Bewertungs- und Kontrollprozess hin untersucht. Im Folgenden werden die Erkenntnisse der Recherche zu relevanten XAI-Anforderungen und der Nutzungskontextanalyse in Anlehnung an das Design Framework von Laato et al. (2022) (Abbildung 3) durch die Beantwortung der Fragen „Wann?“, „Was?“ und „Wie?“ dargestellt.

An dieser Stelle soll erwähnt werden, dass die Analyse für einige der Anforderungsbereiche keine klaren Prioritäten nahegelegt hat. Daher wird die Priorisierung der Anforderungen gemeinsam mit den Funktionalitäten in der Konzeption (Kapitel 4.3) festgelegt.

#### **Wann sollen Erklärungen gegeben werden?**

Erklärungen durch XAI-Systeme sollten in der behördlichen Lebensmittelkontrolle kontextspezifisch und bedarfsorientiert gegeben werden. In der alltäglichen Praxis der Lebensmittelkontrolleure, die eine Vielzahl von Betrieben überwachen, sind Erklärungen besonders dann relevant, wenn ungewöhnliche oder risikobehaftete Situationen auftreten. Beispielsweise können Erklärungen bei der Identifizierung von Betrieben mit hohem Risikopotenzial oder bei der Aufklärung von Regelverstößen hilfreich sein. Da Lebensmittelkontrolleure häufig allein arbeiten, sollte das XAI-System sie dabei unterstützen, Entscheidungen transparent und nachvollziehbar zu treffen, um sowohl die eigene

Entscheidungsqualität sicherzustellen als auch die Kommunikation mit betroffenen Betrieben zu erleichtern.

### **Was sollen die Erklärungen für Informationen enthalten?**

Die Erklärungen eines XAI-Systems in der behördlichen Lebensmittelkontrolle sollten eine Balance zwischen Vollständigkeit und Verständlichkeit halten. Sie müssen relevant, präzise und für den Kontext der Lebensmittelkontrolle angepasst sein. Die Erklärungen sollten die verwendeten Daten und die daraus resultierenden Entscheidungen transparent machen. Dazu gehört die Darstellung der Gründe für eine bestimmte Risikoeinstufung eines Betriebs oder die Erläuterung der Faktoren, die zu einer bestimmten Entscheidung geführt haben. Hierbei ist es entscheidend, dass das System die mentalen Modelle und das Vorwissen der Kontrolleure berücksichtigt, um eine hohe Verständlichkeit und Akzeptanz zu gewährleisten. Gleichzeitig sollten die Erklärungen so gestaltet sein, dass sie auch von Personen mit weniger technischem Hintergrund verstanden werden können.

### **Wie sollen die Erklärungen gegeben werden?**

Die Erklärungen sollten in einer Kombination aus verbaler und nonverbaler Kommunikation erfolgen. Visualisierungen und Metaphern können dabei helfen, komplexe Sachverhalte intuitiv verständlich zu machen. Das XAI-System sollte flexible Erklärungsmöglichkeiten bieten, die sowohl für Routinekontrollen als auch für komplexere Untersuchungen geeignet sind. Dabei ist es wichtig, dass die Erklärungen an die individuellen Bedürfnisse und das Fachwissen der Lebensmittelkontrolleure angepasst sind. Zudem sollten die Erklärungen in die bestehende Softwareumgebung integriert werden können, um eine nahtlose und effiziente Nutzung im Arbeitsalltag zu ermöglichen.

## 4 Konzeption

Im folgenden Kapitel wird die Konzeption des zu implementierenden XAI-Systems detailliert beschrieben. Die Analyse hat wichtige, vorwiegend nicht-funktionale Anforderungen aufgezeigt, die für die Gestaltung eines XAI-Systems im Kontext der Lebensmittelkontrolle berücksichtigt werden müssen. Jedoch stand die finale Bestimmung der funktionalen Anforderungen und der zu implementierenden Funktionalitäten noch aus.

Um dies zu klären, wurden verschiedene Konzeptionsmethoden angewandt. Basierend auf den Ergebnissen der Analyse (Kapitel 3) wurde ein Analyseworkshop (Kapitel 4.1) durchgeführt, um die identifizierten Anforderungen zu strukturieren und Ideen für konkrete Funktionen zu entwickeln.

Ein weiterer Schritt war die Erstellung eines Use-Cases. Dieser diente dazu, ein besseres Bild von der praktischen Anwendung des Systems zu bekommen (Kapitel 4.2). Danach konnten den Anforderungen Funktionen zugewiesen werden (Kapitel 4.3). Um die Spezifikation zu vervollständigen, wurden Annahmen zu dem fiktiven KI-System formuliert (Kapitel 4.4). Anschließend folgt ein iteratives Interfacedesign (Kapitel 4.5).

### 4.1 Analyseworkshop

Im ersten Schritt der Konzeption wurde ein Analyseworkshop durchgeführt. Das Ziel des Workshops war es, die Ergebnisse der Analyse zu strukturieren, Verständnis für das primäre Ziel des Systems zu entwickeln und Ideen für konkrete Funktionen zu entwickeln. Der Workshop sollte außerdem dabei helfen, die eigene Perspektive auf das Thema zu verändern und neue Einsichten zu gewinnen.

#### **Vorgehen**

Teilgenommen haben neben dem Autor (als Moderator) zwei Masterstudierende der Medieninformatik. Im ersten Schritt wurde das Thema der Arbeit, die Zielstellung und die Ergebnisse der Analyse vorgestellt und gemeinsam besprochen. Nachfragen wurden geklärt. Dies stellte sicher, dass alle Beteiligten ein möglichst klares, gemeinsames Verständnis erlangen konnten. Anschließend wurden das Ziel und die Fragestellung des Workshops entwickelt. Im nächsten Schritt hatten die Teilnehmer einige Minuten Zeit, um auf Klebezettel Ideen für Funktionen zu sammeln. Schließlich wurden die vorgeschlagenen Ideen gemeinsam auf dem Whiteboard sortiert und diskutiert. Dies half dabei, die besten Ansätze zu priorisieren.

Die **zentrale Frage** war: Wie können wir Lebensmittelkontrolleuren helfen, eine KI-basierte Risikobewertung von Betrieben nachzuvollziehen?

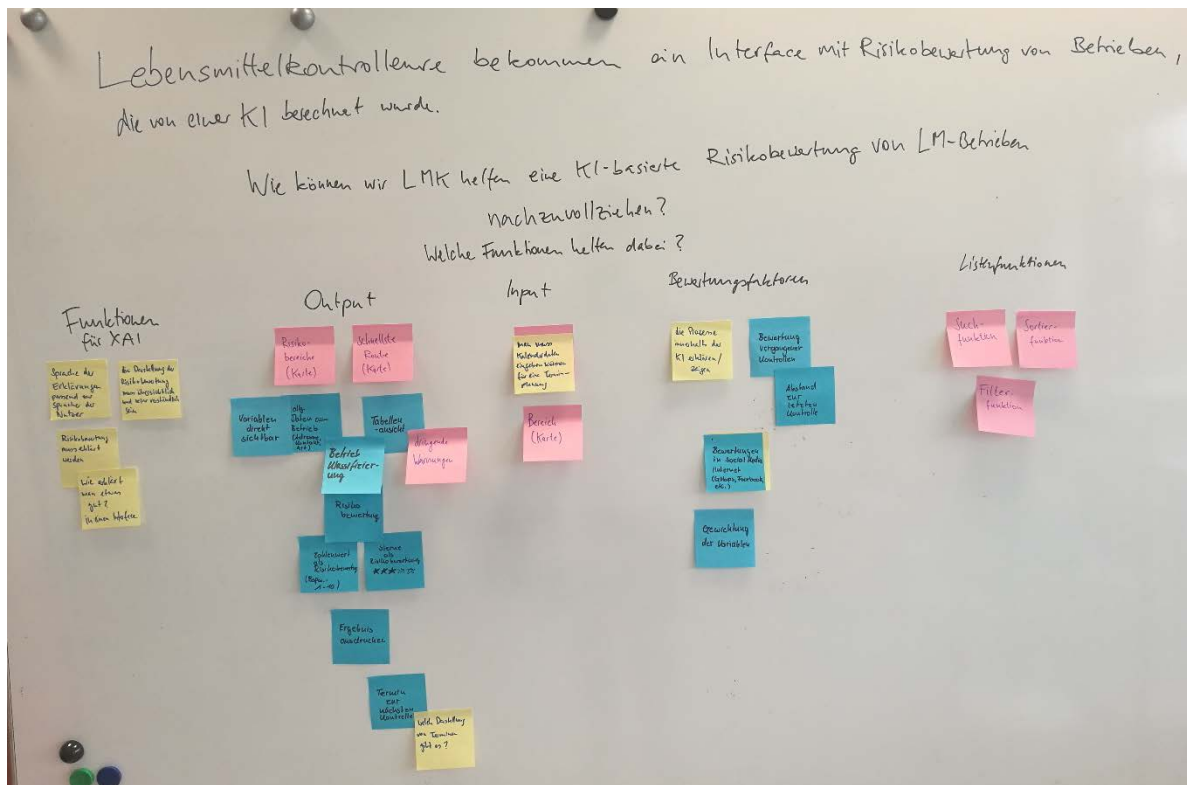


Abbildung 4: Ergebnis des Ideensammelns beim Workshop

## Ergebnis

Als Ergebnis des gemeinsamen Brainstormings ergab sich, dass die Daten und ihre Verwendung in den drei Schritten vor, während und nach der Verarbeitung durch das KI-System zentral für das Verständnis sind: Welche Daten wurden in das System gegeben, wie wurden diese verwendet und interpretiert, und welches Ergebnis wird ausgegeben und wie wird es begründet. Damit liegt die Priorität bei der Gestaltung der Funktionalität auf der Nachvollziehbarkeit der Datenverarbeitung. Die nachfolgende Grafik verdeutlicht den Gedanken.

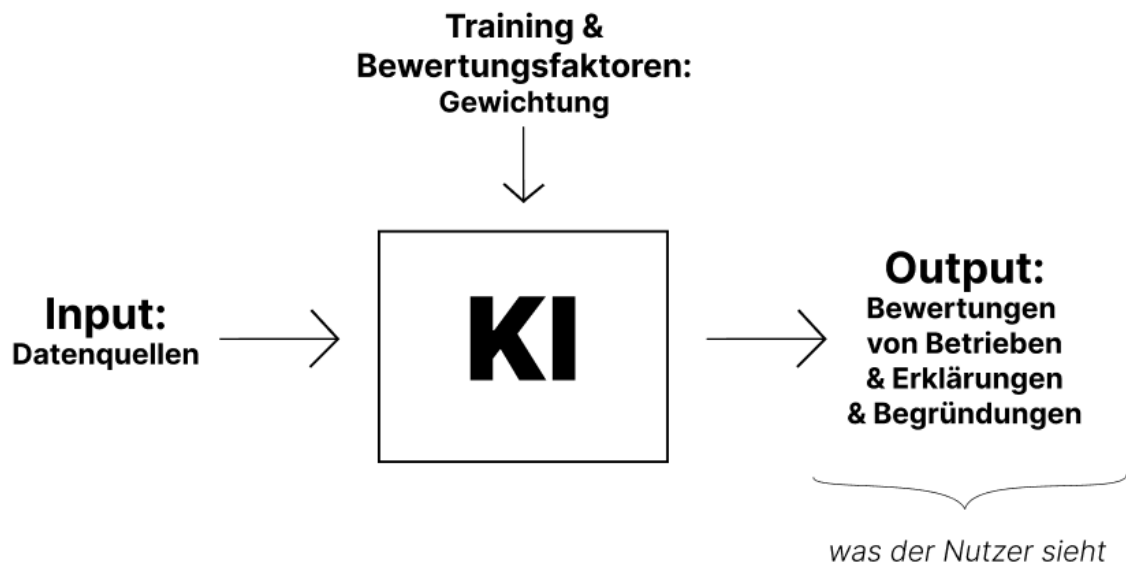


Abbildung 5: Datenverarbeitung des fiktiven KI-Systems

Die sich daraus ergebenden **funktionalen Anforderungen** sind:

- Alle vorhandenen und verwendeten Daten müssen für den Nutzer einsehbar sein
- Das Ergebnis der Berechnung muss deutlich erkennbar und nachvollziehbar begründet sein
- Die internen Prozesse, also die Interpretation bzw. die Gewichtung der Daten, müssen in einem angemessenen Maße vorgestellt werden

Fazit: Die Durchführung des Workshops erwies sich als hilfreich für die Entwicklung von Ideen. Das zentrale Ergebnis des Workshops – die Bedeutung der Nachvollziehbarkeit der Datenverarbeitung – schuf eine solide Basis für die weitere Konzeption.

## 4.2 Use-Case

In der Analyse wurde gezeigt, wie die Kontrollplanung im bisherigen System ohne KI-basiertes Entscheidungsunterstützungssystem abläuft (Kapitel 3.4.1.). Um nun die bisherigen Ergebnisse aus der Analyse und dem Workshop zu nutzen, um einen Arbeitsablauf mit einem XAI-System zu gestalten, wurde ein Use-Case erstellt. Ein Use-Case dient dazu, Szenarien zu beschreiben, in denen ein System eingesetzt wird. Es bietet einen Einblick in die Interaktionen zwischen einem Nutzer und dem System. Das Ziel war es, weitere Einsichten zu benötigten Funktionalitäten zu erhalten. Der folgende Use-Case zeigt daher exemplarisch, wie die Planung von Lebensmittelkontrollen mit einem KI-basierten Entscheidungsunterstützungssystem ablaufen

könnte. Die Beschreibung ist an den Prozess angelehnt, den der interviewte Lebensmittelkontrolleur im selbstgeführten Interview vorstellte (Kapitel 3.4.1).

### **Anwendungsfall für das Planen von Lebensmittelkontrollen**

Claudio Ontrolio arbeitet als Lebensmittelkontrolleur in einer für die Lebensmittelüberwachung zuständigen Behörde im Bundesland Schleswig-Holstein. Zu seinen Aufgaben gehört es, Betriebe, die mit Lebensmitteln zu tun haben, zu kontrollieren. Die Reihenfolge, nach der die Betriebe kontrolliert werden, wird nach einem risikoorientierten Verfahren festgelegt. Dieses Verfahren muss er nicht händisch berechnen. Stattdessen hat er eine Softwarelösung dafür, in die er nur noch einen selbstgewählten Zeitraum eingeben muss und so die zu kontrollierenden Betriebe für diesen Zeitraum von der Software ausgegeben werden.

Die Behörde ist seit Kurzem mit einem KI-basierten Entscheidungsunterstützungssystem ausgestattet. Die Planung der Lebensmittelkontrollen wird nun mit diesem System gemacht. Es ermöglicht eine effizientere Risikobewertung der zu kontrollierenden Betriebe. Damit Claudio also die Möglichkeit hat, mittels der Bewertungen zu entscheiden, welche Betriebe dringend kontrolliert werden müssen, muss er die von dem System ausgegebenen Bewertungen nachvollziehen können.

Es ist Ende des Monats und Claudio verbringt diesen Vormittag im Büro, um verschiedenen administrativen Tätigkeiten zu erledigen. So will er für den nächsten Monat die neue Liste mit den zu kontrollierenden Betrieben erstellen. Er bevorzugt es, eine Liste für einen ganzen Monat zu erstellen, da sich dieser Zeitraum für ihn bewährt hat. Seine Kollegen wählen aber teilweise andere Zeiträume.

Als er zu der Aufgabe der Planung der Kontrollen für den nächsten Monat kommt, öffnet er die neue Software. Nun will er sich die Liste für den nächsten Monat ausgeben lassen. Er trägt den Zeitraum ein, in dem die Kontrollen stattfinden sollen. Nachdem er die Daten eingegeben hat, berechnet das System die im nächsten Monat zu kontrollierenden Betriebe. Während der Berechnung gibt das System als Feedback eine Einschätzung der Berechnungsdauer. Dann gibt das System die Liste mit den Betrieben aus. Sie ist standardmäßig geordnet nach Risikobewertung – von hohem Risiko zum niedrigeren.

Claudio interessiert sich für die Risikobewertung der einzelnen Betriebe. Er will überprüfen, ob die ausgegebenen Ergebnisse korrekt sind. Er öffnet die Erklärung zu der Bewertung des Betriebes mit der höchsten Risikobewertung. Es öffnet sich eine durch das KI-System generierte Zusammenfassung in Textform, die sich Claudio durchliest. Sie klingt zwar plausibel, aber er möchte zur Sicherheit noch in die Rohdaten reinschauen, um sicher gehen zu können, ob das System die Daten wirklich richtig interpretiert hat. Er wechselt zur Ansicht, in der alle vorhandenen Daten dargestellt sind. Da die Rohdaten noch häufig in Papierform vorliegen, werden ihm zuerst nicht die Scans, sondern die wichtigen Auszüge aus den Rohdaten präsentiert. Die Scans der Rohdaten werden ihm aber auch in Miniaturform angezeigt, die er mittels eines Klicks vergrößern und so einsehen kann. Es scheint für ihn, als hätte das System die Daten richtig

interpretiert. Eine Sache ist ihm beim Lesen der Zusammenfassung aber aufgefallen. Das System hat im Text einen bestimmten Punkt besonders hervorgehoben. Claudio geht zur Übersicht der Gewichtung der einzelnen Daten. Ihm fällt auf, dass ein Datenpunkt weniger stark gewichtet wurde, als er angenommen hätte. Er will es genauer wissen und öffnet die Erklärung zur Gewichtung des Datenpunkts und liest sie sich durch. Er ist einverstanden mit der Berechnung und schließt die Ansicht zur Erklärung der Risikobewertung des Betriebs und kehrt damit zurück zur Liste mit den Betrieben vom Anfang.

### 4.3 Funktionalitäten

Wie die Ergebnisse der Analyse und des Workshops, sowie des Use-Case gezeigt haben, teilen sich die Anforderungen und damit die benötigten Funktionalitäten in zwei Aspekte auf:

- Darstellung einer Auswahl der als nächstes zu kontrollierenden Betriebe (Planung der Kontrollen)
- Darstellung von Erklärungen und Begründungen für die Berechnung der Risikobewertungen (XAI)

Bei der Gestaltung des Systems wird ein besonderer Fokus auf die Anforderungen aus Kapitel 3.2 gelegt, um ein System nach XAI-Anforderungen zu gestalten. In der folgenden Tabelle werden die Funktionalitäten den jeweiligen Anforderungen zugeordnet. Eine weitere Spalte zeigt die Priorität.

Anforderung (in Bezug auf)	Funktionalität	Priorität
<b>XAI</b>		
Erklärungen bei Bedarf und Wiederholbarkeit (& Sichtbarkeit der Erklärungen)	Eigener Bereich für die Erklärungen, der sich nach Bedarf öffnen und schließen lässt; das schließt die Sichtbarkeit und einfache Zugänglichkeit mit ein	MUSS
Personalisierte und bedarfsgerechte Erklärungen	Profilbereich mit Einstellungsmöglichkeiten der Erklärtiefe	KANN
angemessene Sprache	Sprache der Erklärtexte werden den Nutzern entsprechend angepasst	KANN

Transparenz der verwendeten Daten und Gewichtung von Faktoren	Daten sind als Rohdaten zugänglich und auch ihre Verwendung wird den Nutzern erklärt	MUSS
Art der Informationsübergabe (Interaktionsform)	Im Rahmen dieser Arbeit reduziert auf größtenteils textliche Erklärungen und Informationen	MUSS
Informationsumfang (& VISUALISIERUNG)	Zusammenfassung der Bewertung ist kurzgehalten, sodass die Hürde die Erklärung zu lesen nicht zu groß ist (& Gewichtung wird als Diagramm dargestellt)	SOLL
CONSIDER what part of the system to explain	Fokus der Erklärungen ist die Verarbeitung der gegebenen Daten; es wird nicht die techn. Seite des KI-Systems erklärt	SOLL
Communicate the uncertainties involved in the Systems decision making	Eine Darstellung der Qualität der Daten bzw. Sicherheit der Berechnung wird dem Nutzer präsentiert	SOLL
<b>Planung der Kontrollen</b>		
Zeitspanne der geplanten Kontrollen	Der Nutzer kann die Daten wählen zum Beginn und Ende des Zeitintervalls, in dem Betriebe kontrolliert werden	MUSS
Ausgabe der zu kontrollierenden Betriebe	Es wird eine Liste mit den zu kontrollierenden Betrieben ausgegeben	MUSS
Darstellung der Betriebsmerkmale und der Risikobewertung	Die erzeugte Liste besteht aus verschiedenen Spalten, u.a. mit der Risikobewertung und verschiedenen Informationen zum Betrieb	MUSS
Kartenansicht	Dem Nutzer werden die zu kontrollierenden Betrieb über eine Kartenansicht dargestellt	KANN

*Tabelle 2: Anforderungen und Funktionalitäten*



**Anmerkung:**

Besondere Herausforderungen stellen die personalisierten und bedarfsgerechten Erklärungen sowie angemessene Sprache, also die inhaltlichen Aspekte der Erklärungen, dar. Diese können nur beispielhaft umgesetzt werden, da es schwer ist, der Realität entsprechende Rohdaten und die darauf aufbauenden Texte für die Erklärungen nachzustellen. Auch die Darstellung der Datenqualität kann nur schemenhaft nachgestellt werden.

## 4.4 Annahmen für das KI-System und Interface

Nachdem in den vorherigen Kapiteln die Anforderungen und Funktionalitäten an das System aufgestellt wurden, musste noch für das fiktive KI-System Annahmen getroffen werden, bevor das Interface gestaltet werden konnte. Wie bereits festgestellt, sollte ein System gestaltet werden, welches vorrangig eine nachvollziehbare Darstellung der Datenverarbeitung bieten soll. Die Nachvollziehbarkeit hängt stark von der inhaltlichen Qualität der Erklärungen ab. Es stellte sich heraus, dass die Entwicklung eines XAI-Systems auf Basis eines fiktiven KI-Systems eine große Herausforderung ist, da die Logik der Erklärungen ein wichtiger Teil der Nachvollziehbarkeit ist und KI-Prozesse in dieser Arbeit möglicherweise nicht befriedigend nachgestellt werden können. Die Annahmen dienen daher als eher grober Rahmen, wie ein für diesen Anlass entworfenes KI-System funktionieren könnte.

Es wird folgendes angenommen:

**Annahme 1:** Das fiktive KI-System erhält als Input verschiedene Datenquellen. Es erfasst sie, es extrahiert Daten aus den Quellen, es interpretiert die Daten und verbindet die Informationen zu einer Bewertung eines Betriebs. Als Output wird eine Liste von nach Risiko bewerteten Betrieben ausgegeben. Der KI-Algorithmus kann seine Berechnungen erklären, indem Verbindungen zwischen Input und Output hergestellt werden.

**Annahme 2:** Es wird angenommen, dass folgenden Datenquellen als möglicher Input gegeben werden:

- bisherige Kontrollprotokolle und daraus resultierende Kontrollabstände
- Daten aus der Gewerbemeldestelle
- Beschwerden aus der Bevölkerung, die direkt an die Behörden getragen wurden
- Bewertungen auf Bewertungsportalen im Internet, wie z.B. Yelp oder Google
- Beschwerden über ein Produkt / Betrieb auf Social Media Plattformen, wie Facebook
- Informationen aus medizinischen Kontexten, also z.B. Krankheiten, die im Umlauf sind
- Informationen über Schädlinge in bestimmten Gebieten

**Annahme 3:** Der von dem KI-System erzeugte Output steht dem zu implementierenden System über eine nicht weiter definierte Schnittstelle zur Verfügung.

**Annahme 4:** Das zu implementierende System wird folgendermaßen in das bestehende behördliche Kontrollsystem integriert:

- zyklische Kontrollen bleiben erhalten, diese werden aber nun zu einer Mischung aus zykl. Kontrollen und Berücksichtigung von akuten Fällen, die zeitnah kontrolliert werden sollten, da es denkbar ist, dass ein KI-System durch Bewertung der aktuellen Datenlage zur Entscheidung kommt, dass ein Betrieb außerhalb seines bisherigen Zyklus kontrolliert werden sollte
  - dafür wird neben dem bisherigen „Zykluswert“ ein zusätzlicher fiktiver Risikowert verwendet. Dieser wird für diese Arbeit als eine Skala von 1 bis 10 definiert.
- das Risiko aller Betriebe wird berechnet nach aktuellem Stand der Daten

**Annahme 5:** Um die Ansicht der Gewichtung der Daten mit Inhalt füllen zu können, musste die Frage wie das KI-System zu dem Ergebnis der Risikobewertung kommt, beantwortet werden. Dazu gehört auch die Frage, wie aus (vorwiegend) qualitativen Daten ein numerischer Skalenwert entsteht. Dafür sind mehrere Möglichkeiten denkbar, diese werden aber nicht weiter spezifiziert.

**Annahme 6:** Das System wird mit einem Büro Setting verwendet, als PC mit externem Bildschirm

## 4.5 Interfacedesign

Im folgenden Abschnitt wird das Vorgehen bei der Gestaltung der Anwendung beschrieben. Es teilt sich auf in zwei Iterationen. Jede der Iterationen wird jeweils mit einer formativen Evaluation in Form von Expertenfeedback abgeschlossen.

### 4.5.1 1. Iteration: Skizzen

In der ersten Iteration wurden Skizzen erstellt, die per Hand gezeichnet und eingescannt wurden. Sie wurden formativ evaluiert, indem sie drei Masterstudierenden der Medieninformatik vorgestellt wurden. Die Skizzen werden im Folgenden nacheinander vorgestellt. In Abbildung 6 wird zunächst die Hauptansicht des Systems dargestellt. Abbildung 7 zeigt die Kartenansicht. Danach folgen in Abbildung 8, 9, 10 und 11 Detailansichten zu den Erklärungen des KI-Systems.

#### **Haupt- bzw. Listenansicht**

In der Skizze werden folgende Funktionen für die Planung der Lebensmittelkontrollen dargestellt:

- Eingabe der Zeitspanne, in der die Kontrollen stattfinden sollen mit jeweils einem Feld für Startdatum und Enddatum
- Liste mit der Risikobewertung der Betriebe (standardmäßig nach Risiko sortiert)

- In jeder Zeile der Liste Buttonelemente zum Öffnen der Erklär-Ansichten
- Sortier-, Filter- und Suchfunktion für die Liste
- Seitenleiste mit den Buttons zum Wechseln zwischen der Listen- und der Kartenansicht
- Kopfleiste mit einem Element, das anzeigt, wann die Daten zuletzt aktualisiert wurden und einem Element für Profilanzeige /-einstellungen (im Profil wird der Bezirk eingestellt)

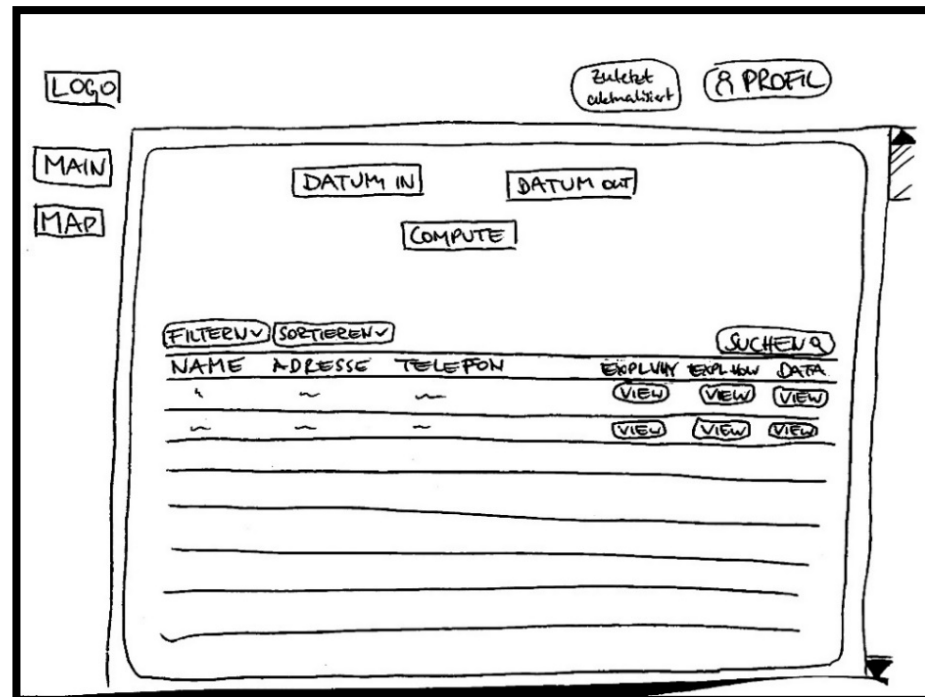


Abbildung 6: Listenansicht

## Kartenansicht

In der Kartenansicht werden die Betriebe auf der Karte angezeigt. Die Betriebe werden zusätzlich an der Seite aufgelistet in einer Klappliste, die weitere Infos zu einem gewählten Betrieb präsentiert.

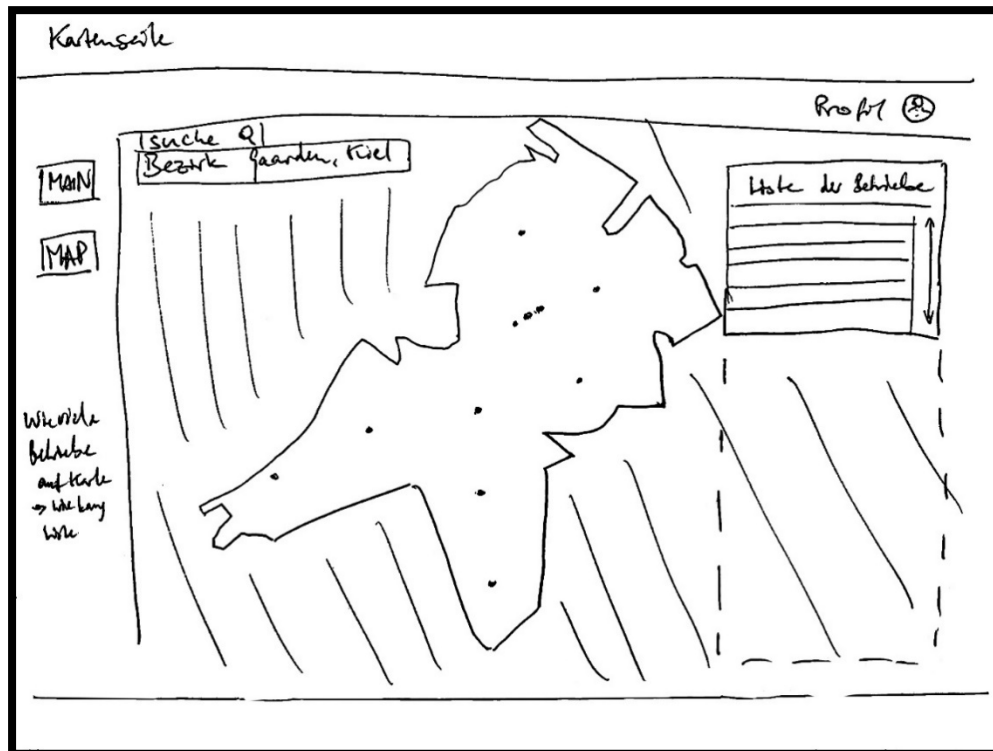


Abbildung 7: Kartenansicht

## Ansichten für die Erklärungen der verschiedenen Aspekte der Datenverarbeitung

### Zusammenfassung

In der Ansicht für die Zusammenfassung wird ein von dem KI-System generierter Text angezeigt, der eine kurze Begründung liefert, warum ein Betrieb seine Bewertung erhalten hat. Zusätzlich werden Visualisierungen von numerischen Daten dargestellt.

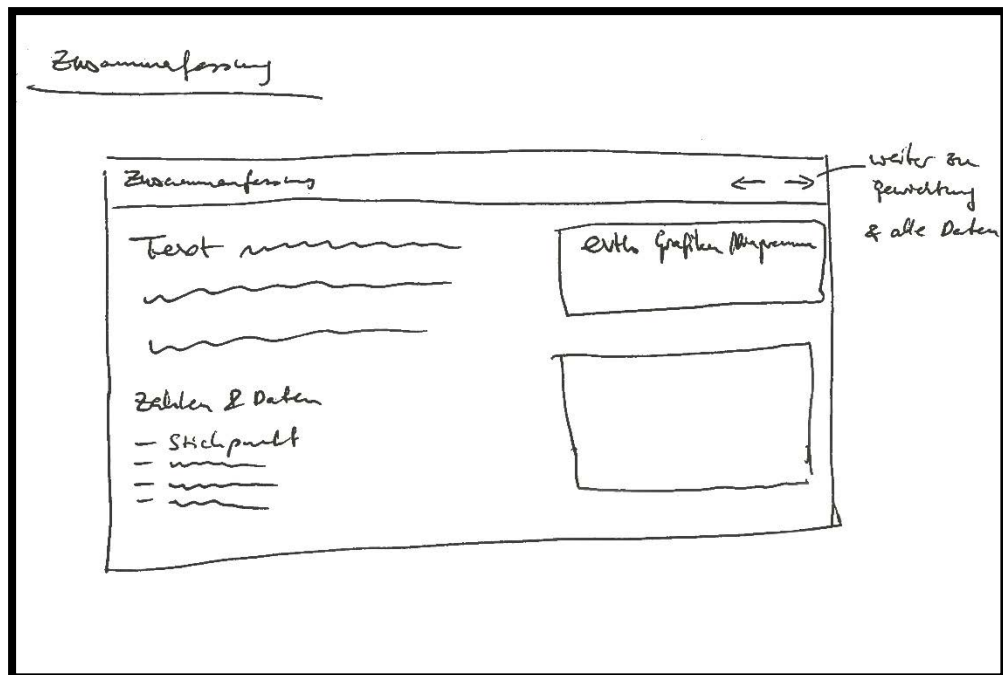


Abbildung 8: Ansicht der Zusammenfassung der Bewertung

## Gewichtung

Die beiden nächsten Ansichten zeigen zwei Möglichkeiten für die Darstellung der Gewichtung der Daten durch das KI-System. Die eine Darstellung bedient sich einem Balken-, die andere einem Tortendiagramm. Die gewichteten Informationen werden daneben in Klapp- oder Scroll-Fenstern angezeigt.

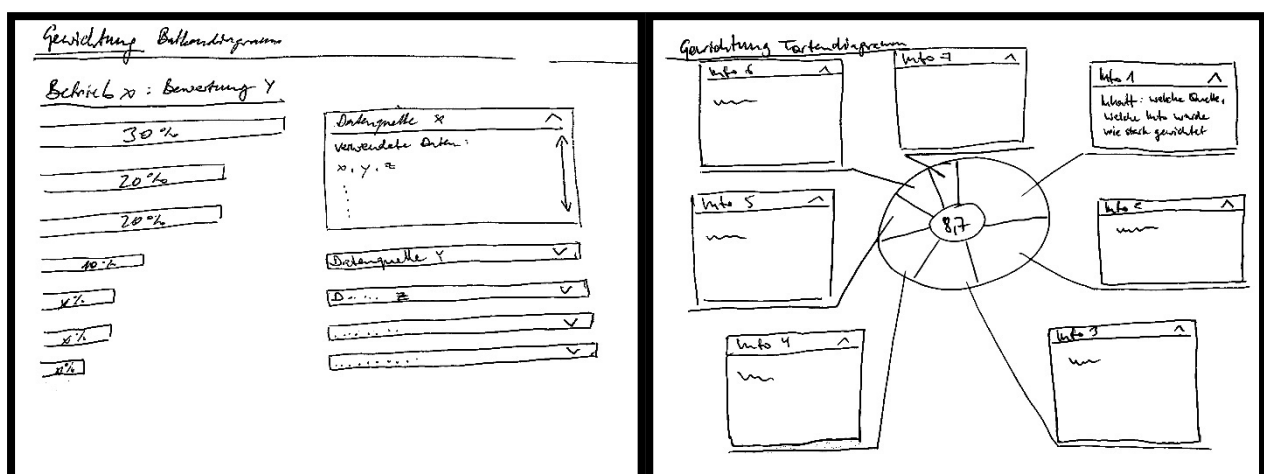


Abbildung 9: Gewichtungsansicht als Balken- & Tortendiagramm

## (Roh)Daten

Die folgende Ansicht gibt dem Nutzer die Möglichkeit die originalen (Roh-)Daten zu betrachten. Jede Datenquelle hat ihren eigenen Bereich, der defaultmäßig eingeklappt ist, sodass beim Öffnen der Ansicht alle Bereiche leicht überblickt werden können. Eine Suchfunktion ergänzt die Seite.

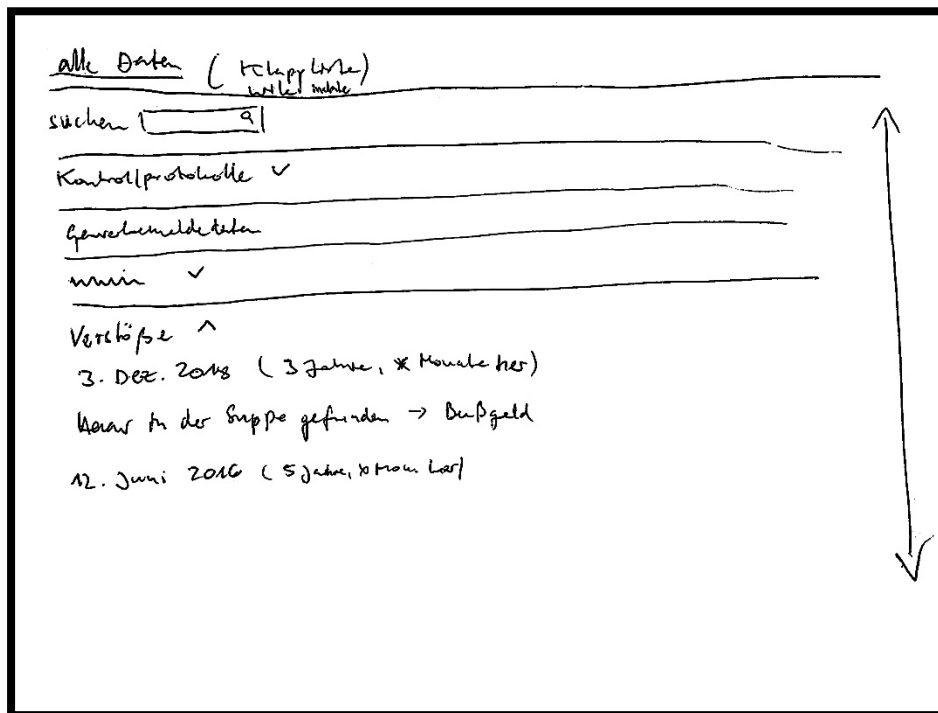


Abbildung 10: Rohdaten

## Ansicht Datenqualität

Die Ansicht zur Datenqualität enthält eine Darstellung der Qualität der vorhandenen Daten zu einem Betrieb. Die Qualität jeder einzelnen Quelle wird in Ampelfarben dargestellt und zu einem Gesamtergebnis zusammengeführt. Da die Bewertung eines Betriebs durch ein KI-System von den vorhandenen Daten abhängt, ist eine transparente Darstellung für die Nutzer wichtig, um die Bewertung nachvollziehen zu können. Die Qualität kann mit Attributen wie Vollständigkeit, Genauigkeit, Zuverlässigkeit und Aktualität beschrieben werden. Die Auswahl der Attribute unterliegt keiner wissenschaftlichen Recherche. Bis zum Schluss der Arbeit konnte nicht geklärt werden, welche Kategorien geeignet sind, um die Qualität zu repräsentieren, weswegen diese im weiteren Verlauf unter Vorbehalt zu betrachten sind.

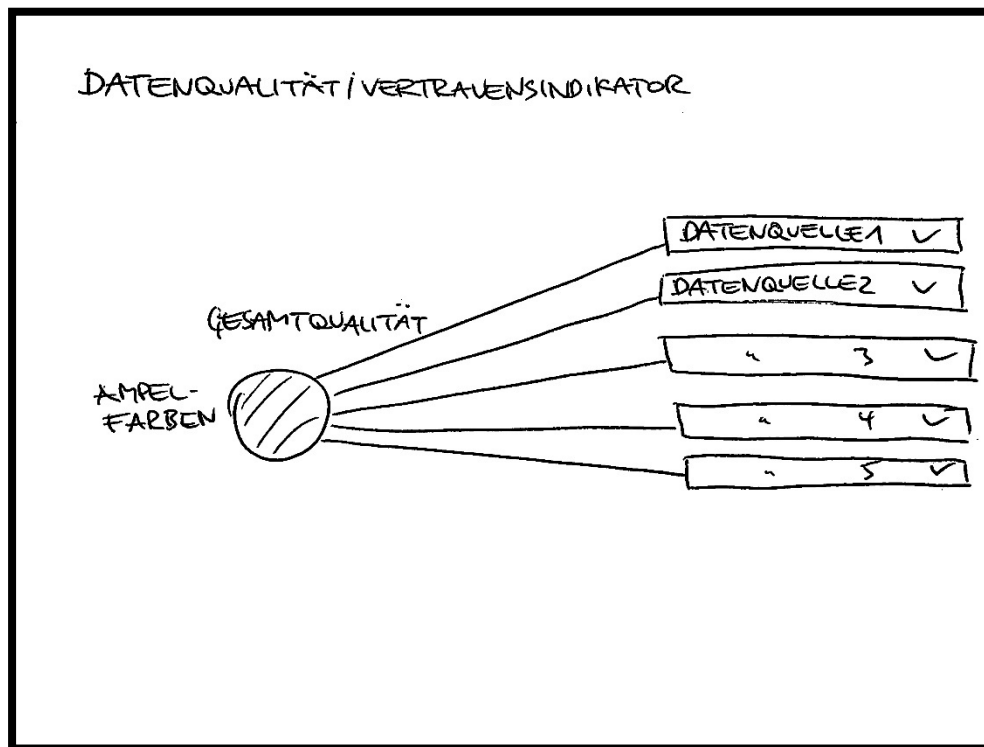


Abbildung 11: Datenqualität

### Feedback zu den Skizzen

Die im vorherigen Abschnitt vorgestellten Entwürfe wurden in einer formativen Evaluation evaluiert. Dieser Abschnitt fasst das Feedback und die Anregungen zu den Skizzen zusammen. Es wurden sowohl positive Elemente als auch Bereiche für mögliche Verbesserungen identifiziert. Die formative Evaluation zielt darauf ab, frühzeitig Erkenntnisse zu gewinnen, um das System iterativ zu optimieren.

Folgende Fragen sollten beantwortet werden:

- Verstehen die Experten das grundlegende Konzept?
- Sind bestimmte Aspekte des Designs verwirrend?
- Wie kann die Gebrauchstauglichkeit verbessert werden?

Da es sich um Skizzen handelt und es stärker um das grundlegende Design und die Funktionalität geht, wurde eine informelle Vorgehensweise gewählt. Zunächst wurden den Teilnehmern der Kontext und die Ziele der Arbeit und des Designs bereitgestellt. Die Teilnehmer wurden gebeten, Fragen zu Unklarheiten jederzeit zu stellen. Dann wurden die Skizzen präsentiert und die einzelnen Komponenten, ihre Funktionen und Ziele erklärt. Danach hatten die Teilnehmer einige

Minuten Zeit, ihre Gedanken mit pro und contra schriftlich zu übermitteln, um die Auswertung zu erleichtern. Anschließend wurden einige der Punkte in gemeinsamer Runde diskutiert.

#### **Auszug der wichtigsten Ergebnisse der formativen Evaluation der Skizzen**

- Vorschläge wurden gemacht, um die Risikobewertung stärker hervorzuheben
- Obwohl die Buttons für die Erklärungsansichten in den Zeilen der Listenansicht als ein schneller und offensichtlicher Zugang zu den Erklärungen gesehen wurde, gab es eine Diskussion über eine mögliche Überlastung der Tabelle durch wiederholte Button-Anzeigen in jeder Zeile. Eine Empfehlung war die Zusammenführung der Erklär-Buttons, um die Anzahl der Klicks zu reduzieren und so die Usability zu erhöhen.
- Die allgemeine Platzierung der Seitenelemente wurde als klar strukturiert bewertet
- Es wurde positiv gewertet, dass die Informationen (in den Detailansichten) von Ansicht zu Ansicht detaillierter werden
- Bei der Wahl zwischen Tortendiagramm und Balkendiagramm für die Darstellung der Gewichtung, wurde das Balkendiagramm generell als besser geeignet empfunden, während das Tortendiagramm aufgrund potenzieller Unübersichtlichkeit kritisiert wurde. Das Balkendiagramm wurde als geordnet und als am einfachsten einzuschätzen bewertet.
- Eine Ansicht aller Daten wurde als sinnvoll gewertet. Es wurde angemerkt, ob eine Anzeige der Datenqualität mitintegriert werden sollte

#### **4.5.2 2. Iteration: Mockups**

Die Skizzen aus der ersten Iteration und die Verbesserungsvorschläge aus dem Feedback wurden im nächsten Schritt in Mid-Fidelity-Mockups umgesetzt. Dafür wurde das methodische Vorgehen des Atomic Design gewählt, um die einzelnen kleinen Elemente zu gestalten, die später zu größeren Komponenten zusammengesetzt wurden. Parallel wurden eine Farbpalette und Schriftarten gewählt. Die Umsetzung erfolgte mit dem Design-Tool Figma (Figma, 2023).

##### **Atomic Design**

Nachdem in der ersten Iteration die Struktur der einzelnen Ansichten des Interfaces konzipiert wurden, folgte im nächsten Schritt die konkrete Gestaltung des Interface und dessen Elemente. Dafür wurde das Vorgehen des Atomic Design gewählt. Bei diesem Vorgehen werden zunächst immer die kleinsten alleinstehenden Elemente eines Designs, die sogenannten Atoms, gestaltet. Dann werden diese einzelnen Elemente nach und nach zusammengesetzt, um neue Elemente des Interfaces zu gestalten (sog. Molecules), wobei diese eine Gruppierung der nächstkleineren Elemente sind. Atomic Design ermöglicht es, Interfaces systematisch und modular aufzubauen, wodurch Konsistenz und Wiederverwendbarkeit der Komponenten gefördert und dadurch die Usability verbessert wird.



Die Umsetzung der Mockups begann mit der Gestaltung der Atoms. In diesem Fall sind das die verwendeten Icons, Buttons, Eingabe- und Textfelder. In Abbildung 12 sind Atoms und die daraus entstehenden Molecules dargestellt.

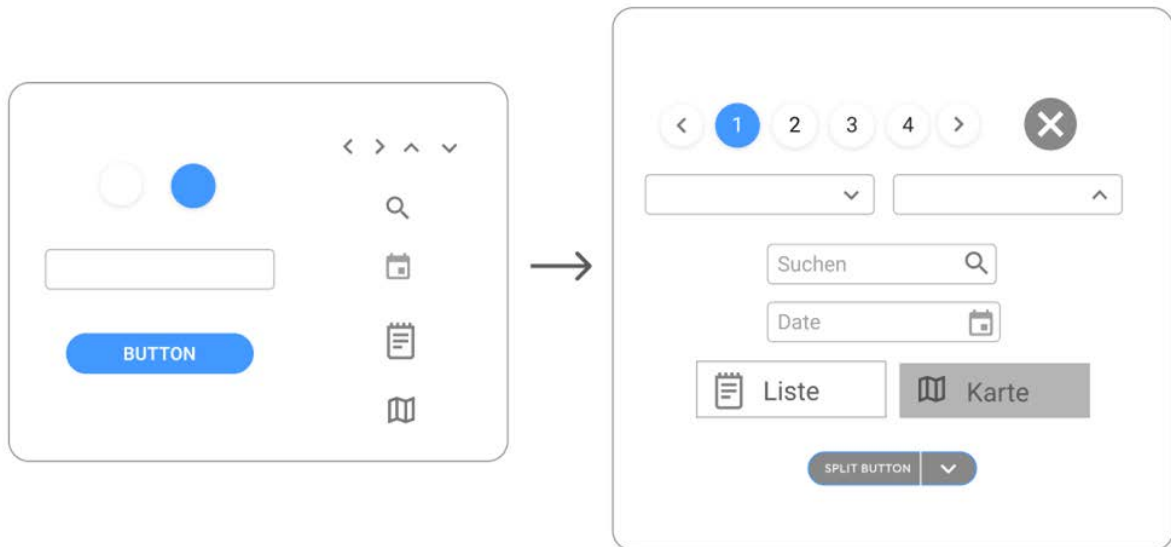


Abbildung 12: Von Atoms zu Molecules

Nachdem die grundlegenden Elemente fertig waren, wurden diese zu den verschiedenen Ansichten zusammengebaut. Dabei entstand die Gestaltung der Kopfleiste und des Eingabebereichs für das Datum. Außerdem wurde das Popup-Fenster mit den Bereichen für die verschiedenen XAI-Ansichten gestaltet. Im letzten Schritt des Atomic-Designprozesses wurden die Elemente zusammengesetzt, um die Gestaltung des kompletten Interfaces abzuschließen.

### Farbgestaltung

Zusätzlich mussten für die Umsetzung der Gestaltung die Farben und die Schrift definiert werden. Da bei der Festlegung des Farbschemas keine anderen Systeme als Vorlage dienten, konnte, unter Beachtung von Usability-Aspekten, frei entschieden werden. Das grundlegende Farbschema (Abbildung 13) orientierte sich an typischen Büroanwendungen mit schlichten Weiß- und Grautönen. Als Akzent- und Hinweisfarbe für Interaktionsmöglichkeit wurde ein Blauton gewählt. Um den Wert der Risikobewertung deutlicher hervorzuheben und dem Nutzer zu signalisieren, ob die Liste nach Risiko sortiert ist oder nicht, wurde der Zahlenwert mit einem farblichen Hintergrund hinterlegt. Dieser besteht aus einer Reihe von 10 Farbtönen beginnend mit einem hellen Gelbton für das niedrigste Risiko bis hin zu einem kräftigen Rotton für das höchste.

Die Gestaltung der Risikobewertung brachte viel Farbe in die Listenansicht. Es wurde angenommen, dass eine farbliche Gestaltung der sich in jeder Zeile wiederholenden Buttons für die Anzeige der Erklärungsansichten in Blau, das Auge farblich überfordert. Deshalb musste ein Kompromiss gefunden werden zwischen der Konsistenz der Verwendung von Blau für

Interaktionselemente (also in diesem Fall Buttons) und der Überfrachtung durch zu viel Farbe. Die Buttons sollten daher mit einem Hover-Effekt versehen werden, sodass die grauen Buttons blau oder blauumrandet werden, sobald der Mauszeiger über sie fährt. Da Mitarbeiter in Behörden in ihren Softwaresystemen geschult werden, kann davon ausgegangen werden, dass diese leichte Inkonsistenz nicht dazu führt, dass diese Funktion übersehen wird. Eine weitere Farbentscheidung musste für die Darstellung der Gewichtung der Faktoren getroffen werden. Hier wurde sich für den Einsatz von abwechselnd hellen und dunklen Graublautönen entschieden.



Abbildung 13: Farben für die Interfacegestaltung



Abbildung 14: Farbspektrum für die Risikobewertung

## Die Mockups

Bei der Gestaltung der Mockups konnten nun die fertigen Einzelteile und die gewählten Farben genutzt werden, um die verschiedenen Ansichten zu erstellen. Zusätzlich sind in der zweiten Iteration die Interaktionsmöglichkeiten zum Wechseln zwischen den verschiedenen Ansichten dazugekommen. Auch das Feedback zu der 1. Iteration ist in die Gestaltung eingeflossen.

## Hauptansichten – Liste und Karte

In der Hauptansicht mit Eingabe der Zeitspanne und der Liste der Betriebe, sind vor allem diese Weiterentwicklungen zu nennen:

- Die Risikobewertung wurde hervorgehoben durch farbige Unterlegung. Verschiedene Versionen wurden gestaltet und bewertet. Die Wahl des Spektrums von Gelb zu Rot

signalisiert, dass auch Betriebe mit einem niedrigen Risikobewertung nicht außer Acht gelassen werden dürfen.

- Statt einzelner Buttons zu den verschiedenen XAI-Ansichten kann nun über einen Splitbutton zu den Ansichten gelangt werden. Die gleichzeitige Anzeige mehrerer Buttons in jeder Zeile wurde als visuell überfordernd eingeschätzt und die Bewertung aller verfügbaren Aktionen stellt eine potentielle kognitive Belastung für den Benutzer dar (Laubheimer, 2019). Eine zeilenweise Wiederholung der Buttons wurde zwar nicht aufgehoben, aber durch Verwendung eines Split-Buttons reduziert. Splitbuttons reduzieren die visuelle Komplexität, indem sie ähnliche Befehle gruppieren. Durch Klicken auf die Beschriftung wird eine Standardaktion ausgewählt, durch Klicken auf den Pfeil wird eine Liste mit allen möglichen Ansichten geöffnet. So ist der direkte Zugang zu den Erklärungen immer noch offen sichtbar, im Gegensatz zu Alternativen, die erst bei einem Klick auf eine Zeile sich öffnen und kollabieren. Wenn die am häufigsten verwendete Option zur Standardoption wird, sinken die Interaktionskosten für die Nutzung dieser Option, da das Menü nicht mehr geöffnet werden muss, um das Element auszuwählen (Laubheimer, 2019)
- Um mehr Platz in der Breite zu bekommen für die Merkmale zu den Betrieben, wurde die Buttons in der Seitenleiste für die Listen- bzw. Kartenansicht zu Tabs am oberen Rand umgestaltet.

The mockup shows a web interface for a restaurant risk assessment system. At the top, there's a header with a logo, a 'zuletzt aktualisiert heute 13:30' status, and a user profile 'Claudio Kontrollotti'. Below the header are tabs for 'Liste' (selected) and 'Karte'. The main area contains a table with columns: Name, Adresse, Telefon, and four 'Spalte' (Column) headers. The 'Risikobewertung' column shows risk scores from 0.9 to 9.5, each with a corresponding colored circle. The 'Explanation' column has a 'View' button with a dropdown arrow. A search bar and filters are at the top of the table. A 'COMPUTE' button is above the table. A dropdown menu is open for the 'View' button of the row with a risk score of 5.9, showing options: 'View Why', 'View How', 'View All Data', and 'View Data Quality'.

Name	Adresse	Telefon	Spalte	Spalte	Spalte	Spalte	Risikobewertung	Explanation
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	9,5	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	8,7	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	7,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	6,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	5,9	View Why View How View All Data View Data Quality
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	4,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	3,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	2,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	1,9	View
Restaurant Zum friesischen Deich	Fischstraße 22 24143	0123456789	EIN WERT	EIN WERT	EIN WERT	EIN WERT	0,9	View

Abbildung 15: Mockup der Listenansicht

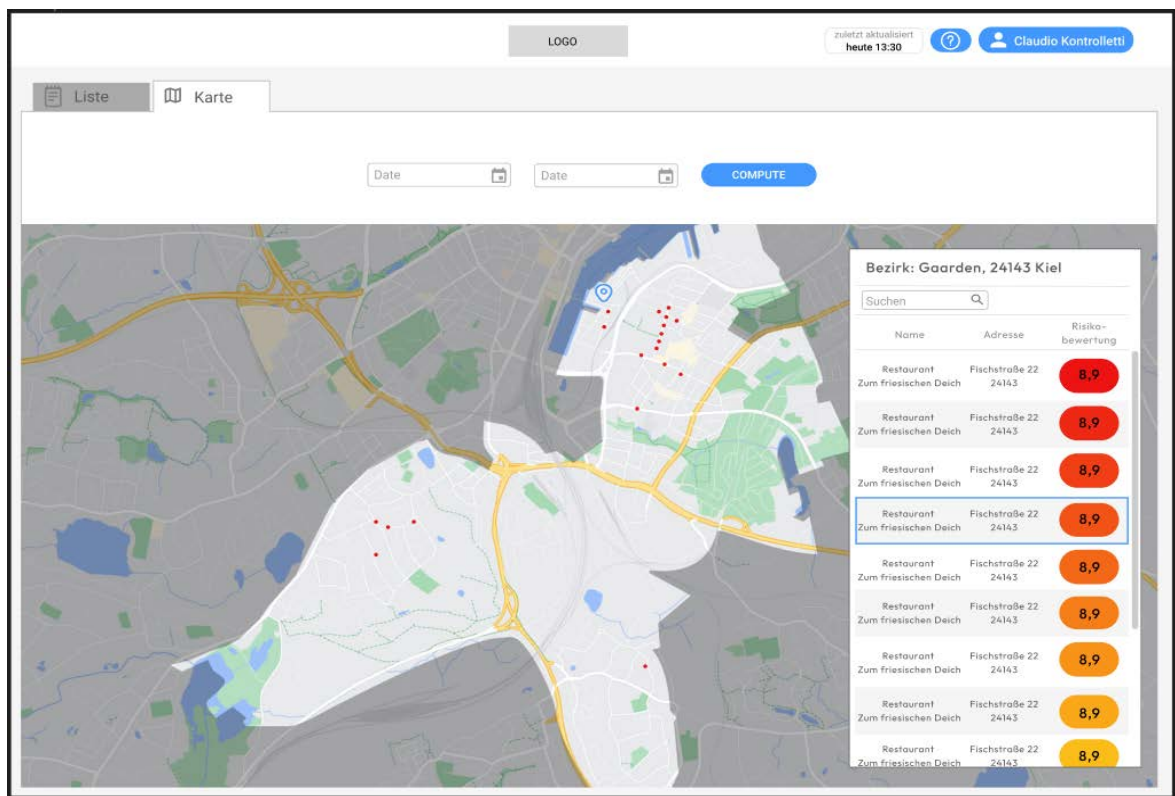


Abbildung 16: Mockup Kartenansicht

## Erkläransichten

Die Weiterentwicklung der Erkläransichten beschränkte sich hauptsächlich auf Farben, Anordnung und Wechsel zwischen den Elementen. Das Wechseln zwischen den Ansichten sollte dem Drill-Down-Prinzip entsprechen. Damit können Benutzer tiefer in Datenebenen oder Funktionen eintauchen, indem sie spezifische Elemente oder Optionen auswählen, was zu detaillierteren Informationen oder spezialisierteren Funktionalitäten führt. Diese Art der Gestaltung fördert eine intuitive Benutzererfahrung, da sie es ermöglicht, schrittweise von einer allgemeinen Übersicht zu spezifischeren Daten vorzudringen, ohne den Nutzer zu überfordern. Weiter sind folgende Details zu den Ansichten hervorzuheben:

- Damit die Nutzer in jeder der Ansichten wissen, um welchen Betrieb es sich gerade handelt, wurde in allen vier Ansichten der Name und die Risikobewertung des Betriebs als Banner eingefügt.
- Der Wechsel zwischen den vier Ansichten erfolgt über ein Pagination-Element. Die farbige Unterlegung signalisiert, auf welcher Seite der Nutzer sich befindet.
- Der Ansicht zu den Rohdaten wurden Miniaturanzeigen für eben diese hinzugefügt. Optimalerweise sind in der Ansicht standardmäßig nicht die einzelnen Dokumente und Quellen als Rohdaten zu sehen, sondern vom KI-System erzeugte Zusammenfassungen der Daten mit möglichen Hervorhebungen wichtiger Stellen. Das ermöglicht es den Nutzern, die Daten leicht zu überblicken. Zusätzlich sollte es den Nutzern aber auch

ermöglicht werden, die Daten in ihrer originalen Form zu betrachten. Dies ist durch die Miniaturansicht mit Vergrößerung möglich.

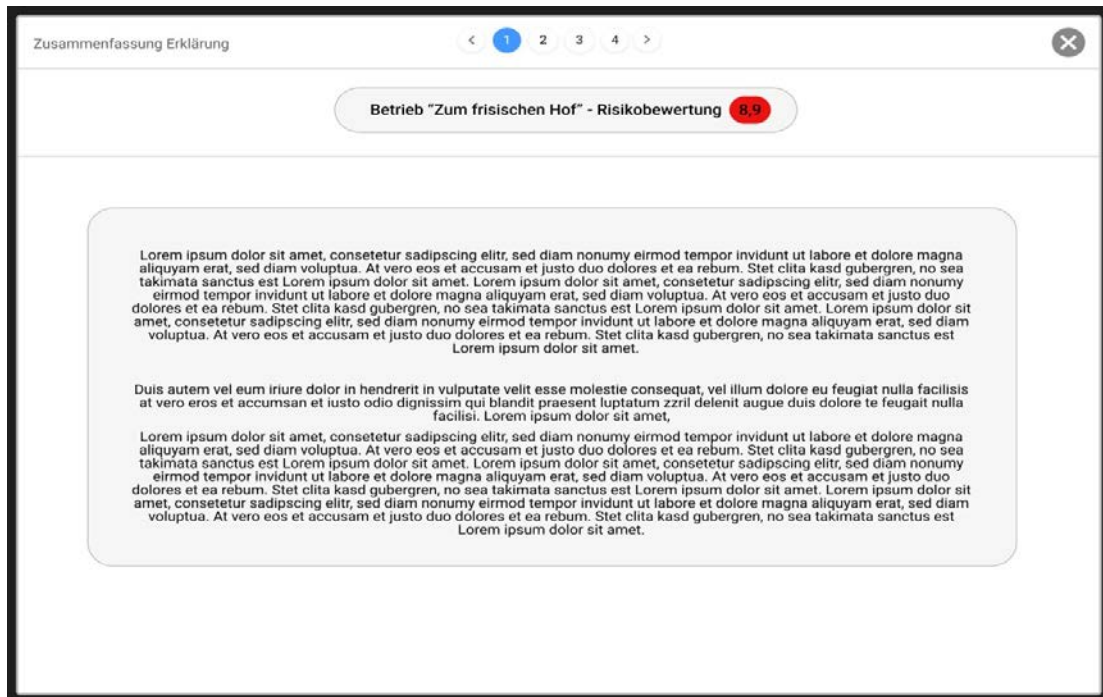


Abbildung 17: Mockup zur Ansicht der Zusammenfassung der Erklärung

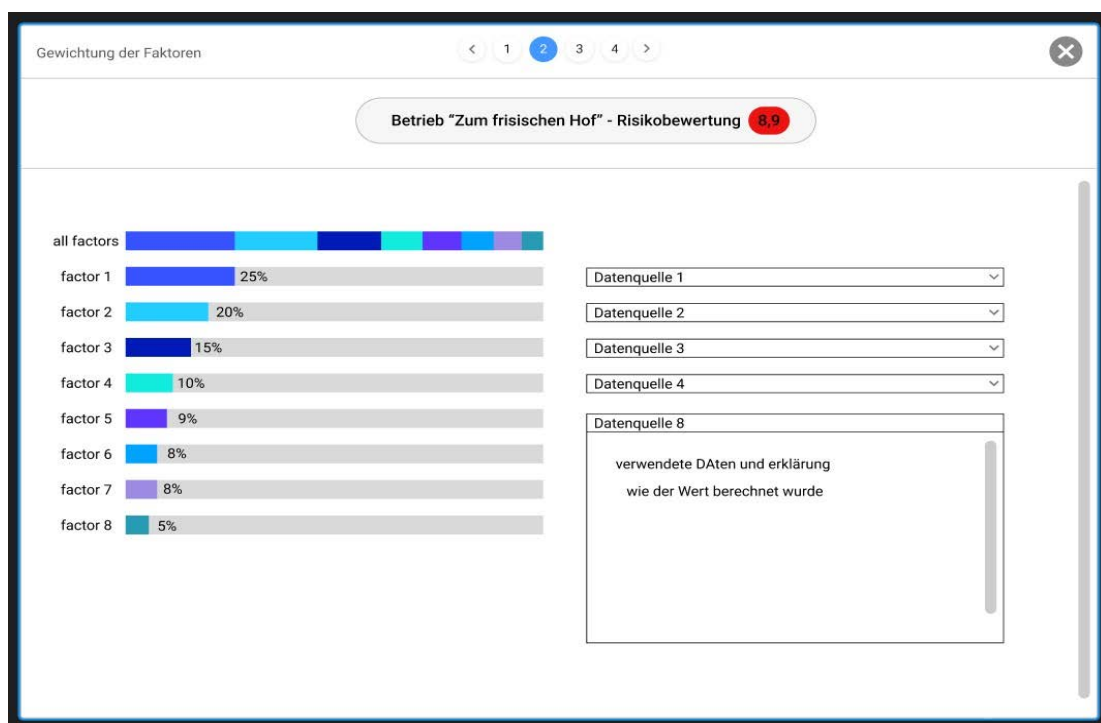


Abbildung 18: Mockup Gewichtungsansicht

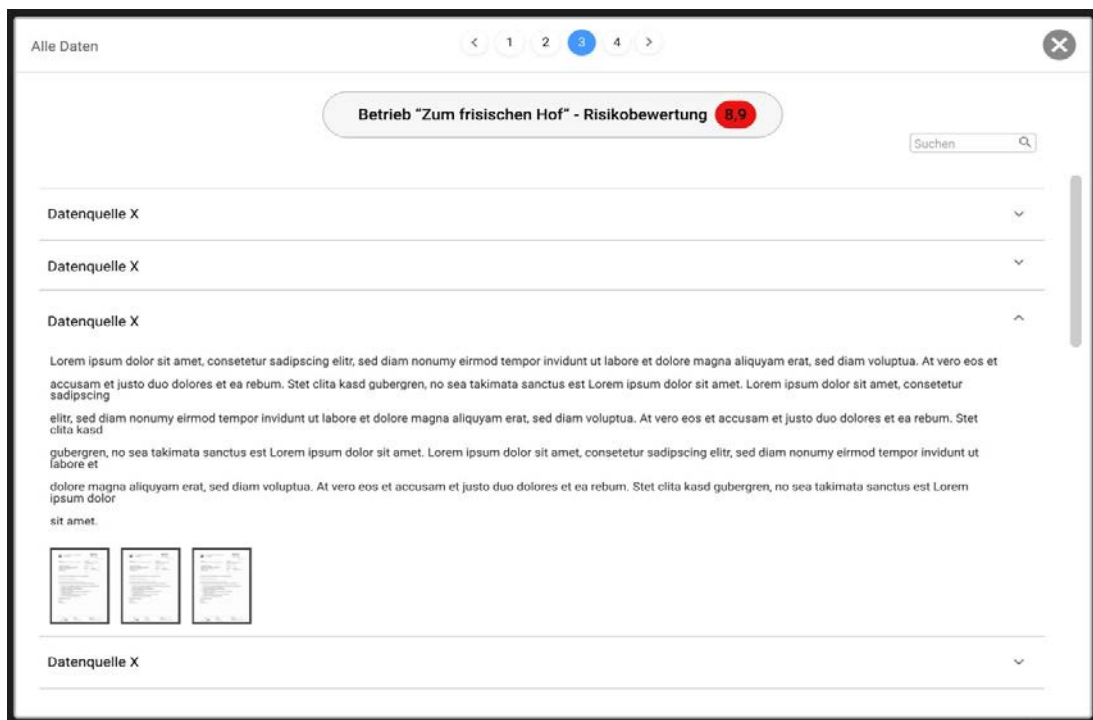


Abbildung 19: Mockup der Rohdatenansicht

### Feedback zu den Mockups

Wie oben gezeigt, wurden aus den initialen Skizzen konkrete Mockups entwickelt, um einer realistischeren und greifbareren Darstellung des Interfaces näherzukommen. Auch die Mockups wurden in einer weiteren formativen Evaluation überprüft, deren Umsetzung und Ergebnisse im folgenden Kapitel dargestellt werden. Das Ziel dieser Evaluation lag darin, die überarbeiteten Designelemente und insbesondere die Interaktionsmöglichkeiten zu evaluieren. Während bei den Skizzen eher ein genereller, übergeordneter Blick auf das Design und die Usability geworfen wurde, liegt der Fokus bei der Evaluation der Mockups verstärkt darauf, wie die Experten die Interaktionselemente zum Wechseln der Ansichten bewerten, ob die hinzugekommene Farbgestaltung verständlich ist und wie intuitiv die Nutzung des Interfaces damit gestaltet ist.

Zu beantwortende Fragen während dieser Evaluation umfassten:

- Sind die Interaktionsflüsse logisch und leicht verständlich?
- Entspricht die Umsetzung den Erwartungen der Experten hinsichtlich der Funktionalitäten, die das System bietet?
- Welche Bereiche oder Elemente des Designs könnten weiterhin Optimierung benötigen?

Die Herangehensweise unterschied sich etwas von der formativen Evaluation der ersten Iteration, um dem fortgeschrittenen Entwicklungsstadium der Mockups gerecht zu werden. Da diese eine höhere Fidelity aufweisen und Interaktionsabläufe fiktiv simulieren können, wurde die

Möglichkeiten eines Paper-Prototyps genutzt, um mögliche Interaktionsschwierigkeiten zu identifizieren.

Im ersten Schritt wurden den Teilnehmern erneut der Kontext und die Ziele der Arbeit und des Designs vorgestellt, jedoch mit Fokus auf den Änderungen, die seit den Skizzen vorgenommen wurden. Nach der Präsentation der Mockups und einer Demonstration der intendierten Interaktionsmöglichkeiten hatten die Teilnehmer die Möglichkeit, mit dem Interface selbstständig zu interagieren. Die Interaktion wurde als Paper-Prototype umgesetzt, in dem die in Figma erstellen Ansichten in Farbe ausgedruckt wurden. Bei der Interaktion mit den Prototypen konnten die Experten, die den Nutzer simulierten, auf ein Interaktionselement auf dem Papier „klicken“, indem mit dem Finger auf das Element getippt oder darauf gezeigt haben. Dann wurde die aktuelle Ansicht mit einer neuen Seite getauscht, die den nächsten Schritt im Interaktionsfluss oder den Zustand der Benutzeroberfläche nach der Aktion repräsentiert. Hierbei konnten die Teilnehmer ihre Gedanken und Erfahrungen während der Nutzung teilen und Fragen stellen. Im Anschluss daran wurden die Teilnehmer gebeten, zu jeder Ansicht positive und negative Aspekte zu notieren. Danach wurde eine Diskussionsrunde durchgeführt, um die verschiedenen Perspektiven und Erfahrungen auszutauschen und gemeinsam mögliche, nächste Schritte zu identifizieren.

### **Auszug der wichtigsten Ergebnisse der formativen Evaluation der Mockups**

#### **Listenansicht:**

- Die Farbgestaltung, die die Risikobewertungen visuell unterstreicht, und das intuitive Design wurden positiv hervorgehoben.
- Kritikpunkte waren u.a. die fehlende Eindeutigkeit bei der Sprachauswahl im UI sowie Optimierungsbedarf bei der Nutzung des Whitespace und der Klarheit von Placeholder-Texten.

#### **Zusammenfassung/Erklärung:**

- Der übersichtliche, zusammenfassende Text zur Risikobewertung wurde positiv aufgenommen.
- Verbesserungsvorschläge betrafen die optische Aufteilung von Text und Wertung zur Vermeidung einer visuellen "Textwand" und eine Größenanpassung des Close-Buttons.

#### **Gewichtung:**

- Die Balkendiagramme wurden als gute visuelle Darstellung der Gewichtung gelobt.
- Verbesserungspotential wurde bei der deutlichen Zuordnung von Faktoren zu Datenquellen gesehen.



### Rohdatenansicht:

- Die klare und geradlinige Gestaltung der "Alle Daten"-Seite wurde positiv hervorgehoben.
- Ein Verbesserungsvorschlag war die Integration einer zusätzlichen Sortierfunktion.

### Datenqualität:

- Die grafische Darstellung wurde für ihre Veranschaulichung gelobt.

## 4.6 Systemarchitektur

Wie bereits an verschiedenen Stellen in dieser Arbeit erwähnt, basiert der in dieser Arbeit erstellte Interfaceprototyp auf mehreren Annahmen zu einem fiktiven KI-System, in dem ein Input von Daten zu Lebensmittelbetrieben zu einer Risikobewertung verarbeitet wurden. An dieser Stelle soll beschrieben werden, wie ein Produktivsystem aussehen könnte und welche Abwandlung für diese Arbeit gemacht wird.

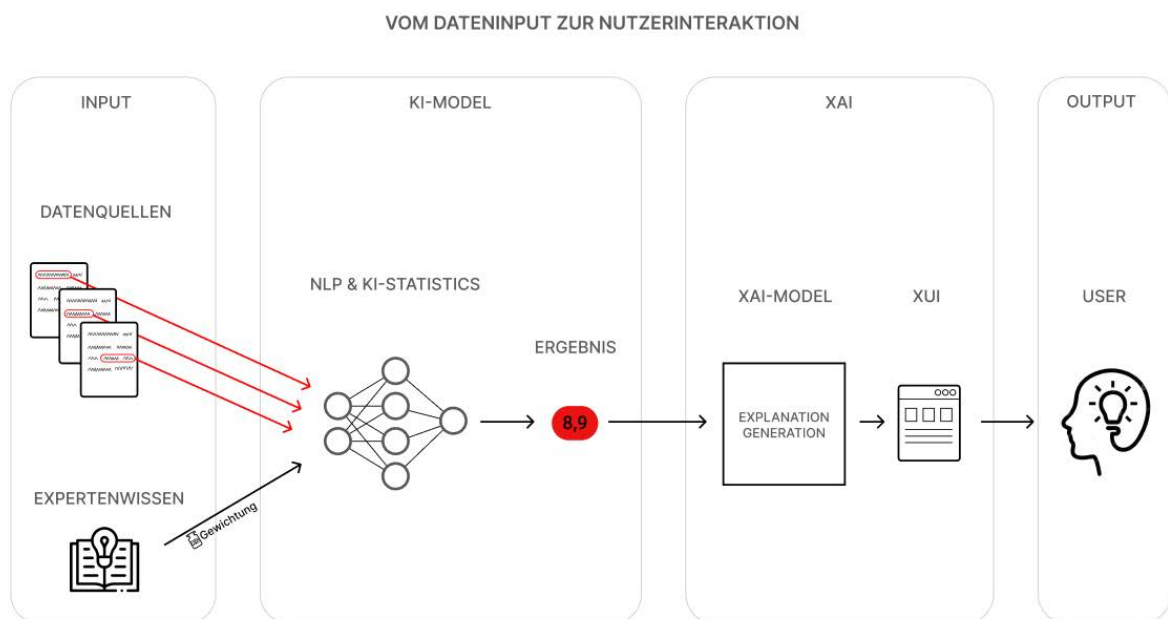


Abbildung 20: Schritte vom Dateninput zum User

Ein in der Praxis anwendbares System würde vier Schritte umfassen:

**Im ersten Schritt** müssten Daten zu den Betrieben vorhanden sein für einen Input. Diese können durch die Kontrollbehörden oder durch computerbasierte Systeme erstellt, bzw. gesammelt sein. Neben Daten zu Betrieben ist vermutlich in diesem Schritt auch Expertenwissen aus der Lebensmittelkontrolle gefragt. Fragen wie:



- Welche Informationen sind relevant?
- Sollten Informationen unterschiedlich gewichtet werden und wenn ja wie?

könnten relevant für ein Produktivsystem sein

**Im zweiten Schritt** werden Daten und Expertenwissen an das KI-Model weitergegeben. Das KI-Modell kann mit verschiedenen Verfahren den Input zu dem gewünschten Ergebnis verarbeiten. Denkbare Verfahren wären:

- Textanalyse mit Natural Language Processing (NLP). NLP ermöglicht es, Textdaten zu verstehen und daraus Schlussfolgerungen zu ziehen.
- Identifizierung von Mustern und Auffälligkeiten, also das Analysieren von Daten, um wiederkehrende Themen und gemeinsame Probleme zu identifizieren.
- Statistische Zusammenhänge aus größeren Datenbeständen zu identifizieren, um Wahrscheinlichkeiten für auftretende Ereignisse zu gewinnen und so eine Art Forecasting-System zu erhalten.

Diese Verfahren in Kombination mit dem Einbinden von Expertenwissen ermöglicht die Berechnung der Risikobewertung eines Betriebes.

**Im dritten Schritt** wird aus der Berechnung des KI-Modells eine Erklärung mittels des XAI-Modells generiert und diese im XUI (Explainable User Interface) dargestellt. Mögliche Verfahren zur Generierung von Erklärungen sind:

- Natural Language Explanations: Einige Modelle fokussieren sich darauf, Erklärungen in natürlicher Sprache zu generieren, um menschenähnliche Erklärungen für die Vorhersagen und Entscheidungen des Modells zu liefern. Dabei kann ein Modell trainiert werden, um Erklärungen zu generieren, die mit den Entscheidungen des Modells kohärent sind.
- LIME (Local Interpretable Model-Agnostic Explanations): LIME zielt darauf ab, die Vorhersagen eines maschinellen Lernmodells in einer für Menschen nachvollziehbaren Form zu erklären. Es tut dies, indem es sich auf einen einzelnen Datenpunkt konzentriert und versucht zu ergründen, wie das Modell seine Vorhersage für diesen speziellen Punkt trifft.
- Attention Mechanisms (bei neuronalen Netzen): Zeigt, welchen Teil der Eingabedaten das Netz für eine bestimmte Vorhersage als besonders wichtig erachtet. Das kann helfen, zu visualisieren und zu verstehen, auf welche Aspekte die KI besonders achtet.

Sowohl das KI-Modell als auch das XAI-Modell würden ihre Berechnungen serverseitig durchführen und somit als Backend fungieren. Die errechneten Ergebnisse des XAI-Modells würden dann an einen Client, dem Explainable User Interface (XUI), weitergegeben.

**Im letzten Schritt** betrachtet der Nutzer die erzeugten Erklärungen. Wünschenswert ist hier, dass die Nutzer in das XUI-System eingearbeitet wurden, sodass sie sich auf den Inhalt der Erklärungen konzentrieren können und nicht durch das Interface abgelenkt werden.

Für diese Arbeit wird lediglich das XUI umgesetzt. Die Daten für den Input sind selbsterstellte Dummy-Daten in Form von natürlichsprachlichen Texten und Kartengrafiken. Das Interface besteht aus einer Hauptansicht, einer Kartenansicht und 4 verschiedenen Erkläransichten, die als Popup in der Hauptansicht geöffnet werden können. Die Herausforderung für diese Arbeit ist es, das Interface mittels eines modernen Frontend-Frameworks als Single-Page-Application zu implementieren (siehe Kapitel 5).

## 4.7 Fazit

In der Konzeption wurde mittels verschiedener Methoden und einer iterativen Vorgehensweise ein System gestaltet, das zwei Hauptfunktionalitäten bieten sollte: die Darstellung der Erklärungen und Begründungen von KI-generierten Risikobewertungen (XAI) und die Planung von Lebensmittelkontrollen.

Im Verlauf der Konzeptionsphase hat sich herausgestellt, dass sich die Prioritäten der Arbeit stärker Richtung Primärziel verschoben haben. Das in Kapitel 1.2 genannte Sekundärziel der Entwicklung einer Gestaltungslösung, die die Nutzer bei der Aufgabe Betriebskontrollenplanung unterstützt, wurde bezüglich ihrer Funktionen eingeschränkt. Dies resultiert hauptsächlich aus dem unvorhergesehenen Mehraufwand bei der Erstellung von Datensätzen für die XAI-Funktionalität des gestalteten Systems. Ohne diese Datensätze ist eine Bewertung des Systems nicht möglich. Zudem ließen sich aus der Nutzungskontextanalyse nicht genügend Anforderungen für eine KI-gestützte Planung von Lebensmittelkontrollen ableiten. Es könnte nur darüber spekuliert werden, welche Funktionen für eine dynamische Liste erforderlich wären, wie es sie bei einer KI-basierten Listenerstellung gäbe.

Die Funktionalität der Kontrollplanung besteht nun darin, dass den Nutzern eine Liste mit den als nächstes zu kontrollierenden Betrieben präsentiert wird. Neben der Listeansicht, die verschiedene Informationen und die Risikobewertung zum Betrieb enthält, lassen sich die Betriebe auch in einer Kartenansicht betrachten.

Die XAI-Funktionalität bietet verschiedene Ansichten für Erklärungen und Begründungen, wieso die Betriebe ihre Risikobewertung erhalten haben. Es konnten viele Anforderungen aus der Analyse als Funktionalität in das entworfene Konzept integriert werden. Besonders die iterative Vorgehensweise erwies sich als hilfreich bei der Gestaltung des Systems. Durch eine formative Evaluation in jeder Iteration konnten die Verbesserungsvorschläge aus dem Feedback in die nächste Runde integriert werden, um so das Konzept zu verbessern.

## 5 Realisierung

Dieses Kapitel beschreibt die Realisierung des entworfenen Systems. Im Kapitel Konzeption wurde anhand der Analyse ein Konzept erarbeitet. Nun wird das Vorgehen bei der Implementierung dieses Konzepts erläutert. Zuerst werden in Kapitel 5.1 die für die Entwicklung geeigneten Werkzeuge vorgestellt. In Kapitel 5.2 folgt eine Darstellung der Code-Struktur. Danach wird in Kapitel 5.3 die Umsetzung einzelner, aber wichtiger Funktionalitäten genauer betrachtet. Abschließend folgt in Kapitel 5.4 das Fazit zur Implementierung.

### 5.1 Verwendete Werkzeuge

#### **Vue.js**

Die Auswahl an JavaScript-Frameworks für die Entwicklung von Single-Page-Webanwendungen ist vielseitig. In dieser Arbeit wurde Vue.js verwendet. Vue.js basiert auf den grundlegenden Technologien der Webentwicklung: HTML, CSS und JavaScript. Es zielt auf die Erstellung vielfältiger webbasierter Frontend-Interfaces ab. Charakteristisch sind dabei die Vue-Komponenten, die dynamische Elemente des Interfaces darstellen. Diese Komponenten werden in Vue-Dateien definiert und können dann dynamisch von anderen Komponenten geladen werden. Ein Vue.js-basiertes Interface besteht aus mehreren dieser Komponenten. Eine Vue-Datei setzt sich aus drei Bereichen zusammen: „Template“, „Script“ und „Style“. Im „Template“-Bereich wird der strukturelle Aufbau der Komponente mit HTML festgelegt, im „Script“-Bereich lassen sich Funktionen mit JavaScript umsetzen, und der „Style“-Bereich enthält CSS für das Design der Komponente. Ein weiterer Pluspunkt von Vue.js ist neben seinem auf Komponenten basierendem Aufbau die einsteigerfreundliche Lernkurve, der Community-Support sowie ein umfangreiches Ökosystem mit zahlreichen Plug-Ins und Packages.

#### **Bootstrap**

Bootstrap ist ein weit verbreitetes Open-Source-Frontend-Framework, das für die Entwicklung mit HTML, CSS und JavaScript entwickelt wurde. Mit Bootstrap werden Webseiten und Webanwendungen gestaltet und entwickelt, indem eine Sammlung vordefinierter HTML-, CSS- und JavaScript-Komponenten zur Verfügung steht, die das Design und die Funktionalität vereinfachen und schnelle Entwicklung ermöglichen (Bootstrap).

#### **Chart.js**

Chart.js ist eine Open-Source Javascript Bibliothek, die es ermöglicht Daten zu visualisieren. Die Visualisierung kann mittels verschiedener Arten von Diagrammen erfolgen, u.a. Balkendiagramme. Die Diagramme werden als HTML5 Canvas gerendert. Dieses Canvas-Element stellt einen Container bereit, in dem mithilfe von Javascript die Grafik erstellt wird. Chart.js bietet

zudem verschiedene Möglichkeiten, um die erstellten Diagramme anzupassen und zu gestalten (Chart.js).

### **Konva.js**

Für die Implementierung der Datenqualitätsanzeige wurde Konva.js genutzt. Konva.js ist eine JavaScript-Bibliothek zur Erstellung von 2D-Canvas-Rendering. Sie ermöglicht Entwicklern, grafische Benutzeroberflächen und andere visuelle Elemente auf Webseiten und Webanwendungen zu erstellen und zu verwalten. Konva.js zeichnet sich durch eine einfache und intuitive API aus, die es auch weniger erfahrenen Entwicklern ermöglicht, komplexe Grafiken und Interaktionen zu implementieren. Weitere Vorteile sind eine hohe Kompatibilität mit verschiedenen Browsern und die Unterstützung für Eventhandling und Animationen. (Konva.js)

## 5.3 Struktur des Projekts

Die Datei `main.js` bildet den Ausgangspunkt der Codestruktur. Hier wird die Komponente `App.vue` importiert und die Vue-Instanz erstellt. `Main.js` integriert wichtige Abhängigkeiten und Bibliotheken wie den Vue-Router für das Routing und Bootstrap für das Styling. Zudem wird eine Sammlung von FontAwesome-Icons eingebunden und in der Bibliothek zur Verwendung in der gesamten Anwendung verfügbar gemacht.

Die Komponente `App.vue` agiert als zentrales Element der Anwendung. Ihr Hauptzweck besteht darin, die verschiedenen Ansichten der App zu rendern, was durch das `router-view`-Element ermöglicht wird. Dies erlaubt eine dynamische Darstellung der Inhalte basierend auf der aktuellen Route.

Die Navigation innerhalb der Anwendung wird durch die `routes/index.js`-Datei festgelegt. Hier sind verschiedene Routen definiert, die auf spezifische Komponenten verweisen. Es gibt zwei Hauptansichten, `ListView.vue` und `MapView.vue`, die unter den Routen `/list` und `/map` zugänglich sind. Darüber hinaus existiert eine Standardroute, die den Benutzer automatisch auf die Listenansicht umleitet, falls keine spezifische Route angegeben wird.

Die `List.vue`-Komponente, als Hauptansicht der Anwendung, beinhaltet weitere Unterkomponenten wie `NavBar`, `TabButton`, `DateNCompute` und `Table`, die spezifische Funktionen und Darstellungen innerhalb der Listenansicht ermöglichen.

Die darin enthaltene `Table.vue`-Komponente ist ein weiteres zentrales Element. Sie ist verantwortlich für die Darstellung der Daten in Tabellenform. Diese Komponente integriert Unterkomponenten wie `SearchForm`, `RiskScore`, `SplitButtonAlt` und `PopUp`. Die `PopUp.vue`-Komponente enthält mehrere Unterseiten, die die unterschiedlichen Informationsaspekte des XAI-Systems darstellen.

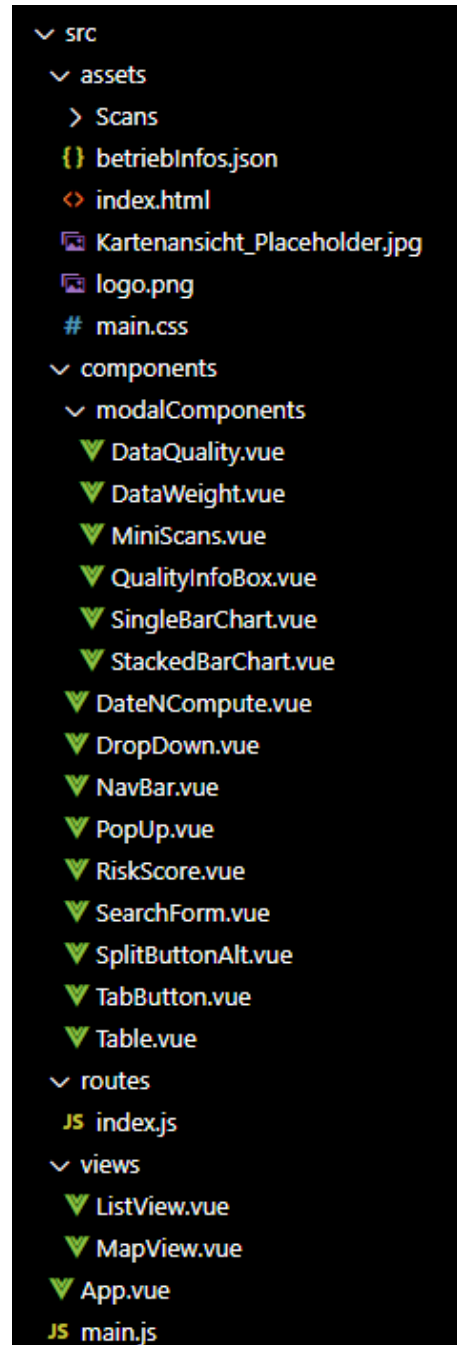


Abbildung 21: Code-Struktur

## 5.4 Implementierung der Funktionalitäten

In diesem Abschnitt wird die Implementierung ausgewählter Funktionalitäten vorgestellt.

### 5.4.1 Umsetzung der Content Verwaltung

Die Verwaltung der verschiedenen Inhalte an den verschiedenen Stellen des Interfaces ist eine zentrale Aufgabe des Systems. Im folgenden Abschnitt werden die einzelnen Teile, die an dieser Aufgabe beteiligt sind, dargestellt.

#### PopUp.vue

Die PopUp.vue Komponente verwendet Daten, die aus einer JSON-Datei stammen. Über diese Komponente werden die Inhalte an die verschiedenen Ansichten in dem PopUp-Fenster verteilt. Um zu verstehen, wie diese Daten verwaltet und verteilt werden, ist es hilfreich, den Prozess in mehrere Schritte zu unterteilen:

1. Import der Daten aus einem JSON-File: Die JSON-Daten werden am Anfang des Script-Teils der Komponente aus der Datei `betriebInfos.json` importiert.

```
import betriebInfos from "../assets/betriebInfos.json";
```

Abbildung 22: Import der JSON-Datei

2. Speichern der Daten in der Komponente: Die importierten Daten werden in der `data()`-Funktion der Komponente gespeichert. Dies ermöglicht es der Komponente, auf die Daten zuzugreifen und sie innerhalb des Template-Teils zu verwenden.

```
data() {  
  return {  
    entries: betriebInfos,  
  };  
},
```

Abbildung 23: entries Variable

3. Verfügbarmachen der Daten für andere (Unter-)Komponenten: Eine Methode namens `getEntryById()` wird verwendet, um auf spezifische Einträge in den Daten zuzugreifen, basierend auf einer ID. Die Methode sucht in den `entries` nach der entsprechenden ID und gibt den spezifischen Eintrag zurück:

```

getEntryById(id) {
  const result = this.entries.find((entry) => entry.ID === id);
  console.log("id:", id, "result:", result);
  return result;
},

```

Abbildung 24: Methode zum Filtern des JSON Inhalts

Zum Überprüfen der Methode wurde eine Anweisung für eine Konsolenausgabe geschrieben, die die ID und den Eintrag ausgibt.

4. Weitergabe der Daten an die Unterkomponenten: Die Weitergabe der Einträge an die verschiedenen Komponenten für die Erkläransichten erfolgt durch das Binden von Props an die Unterkomponenten. Beispielsweise für die Datenqualitätsansicht:

```

<DataQuality :content="getEntryById(contentID)" v-if="page === 4" />

```

Abbildung 25: Weitergabe von Daten an Unterkomponenten

Die Unterkomponenten erhalten die Daten dann als Props und können diese weiterverwenden.

Für die Entscheidung, welcher Inhalt an welcher Stelle im Interface bereitgestellt werden muss, sind hauptsächlich zwei Props verantwortlich:

Mit der contentID wird der richtige Eintrag für die verschiedenen Informationen zum Betrieb aus der JSON-Datei geholt. Die contentView-Prop gibt an, welche der vier Ansichten erwartet wird. Dafür wird sie in der data()-Funktion von PopUP.vue als page gespeichert.

```

props: {
  contentView: Number,
  contentID: Number,
},

```

Abbildung 26: content Props

## Verwendung von JSON

Da die implementierte Anwendung eine große Menge Daten beinhaltet und es für die Funktionalität des Prototyps nicht relevant war, in welcher Form die Daten verfügbar sind, wurde eine einfache und schnelle Lösung gesucht, um die Daten lokal zu speichern und direkt zugreifbar zu machen, ohne auf eine komplexe Datenbankstruktur angewiesen zu sein. Die Wahl fiel auf das JSON-Format.

JSON (JavaScript Object Notation) ist ein textbasiertes Datenformat, das hauptsächlich für den Datenaustausch zwischen einem Server und einer Webanwendung verwendet wird. Es ist für Menschen leicht lesbar und für Maschinen einfach zu parsen und zu generieren. Einer der Hauptvorteile von JSON liegt in seiner Einfachheit und Flexibilität: Es kann komplexe

Datenstrukturen wie Objekte und Arrays repräsentieren und reduziert gleichzeitig die Komplexität des Systems, da keine Kenntnisse im Datenbankmanagement oder in der Konfiguration von Datenbankservern erforderlich sind.

```
assets > {} betriebInfos.json > ...
[
  {
    "ID": "01",
    "Betrieb": "Fischhandel Meeresbrise",
    "Adresse": "Bremer Str. 45, 24143 Kiel",
    "Telefon": "0431 12345678",
    "Ansprechperson": "Otto Hansen",
    "Spalte1": "Ein Wert",
    "Spalte2": "Ein Wert",
    "Spalte3": "Ein Wert",
    "Value": "6.5",
    "Summary": "<div><p>Die Analyse für den \"Fischhandel Meeresbrise\" integriert ve",
    "Gewichtung": [23.33, 23.33, 23.33, 20, 10],
    "Labels": ["Kontrollbericht vom 22.03.2022", "Kontrollbericht vom 22.04.2023", "K",
    "DatensatzGesamt": {
      "Betrieb": "Fischhandel Meeresbriese",
      "buttonText": "Übersicht der Datengewichtung",
      "GesamtRisikoWert": "6.5",
      "EinzelRisikoWert": ["9.1", "6.3", "6.9", "6.2", "5.1"],
      "dropdownText":
        {
          "Text1": "Die Untersuchung zeigte deutlich, dass die fortwährenden und sorg",
          "Text2": "3 Betriebskontrollen in den letzten 12 Monaten",
          "Text3": "mehrere Rezensionen von Bewertungsportalen und Social Media",
          "Text4": "Gewerbeanmeldedaten"
```

Abbildung 27: Ausschnitt des Datensatzes in der JSON-Datei



## 5.4.2 Umsetzung des Interfaces

### Farbindikator für die Risikobewertung

RiskScore.vue implementiert den Farbhintergrund für den Risikowert. Die Component ist so konzipiert, dass sie einen numerischen Wert value als Eingabe über props erhält und basierend auf diesem Wert die Hintergrundfarbe eines mit CSS gestylten Div-Elements ändert.

Im Script-Teil wird das value-Prop definiert. Es gibt eine computed-Eigenschaft color, die den Hintergrundfarbenwert basierend auf dem Wert von value liefert. Die Farben ändern sich in einer spezifischen Reihenfolge von Gelbtönen zu Rottönen.

Die computed-Eigenschaft für die dynamische Zuweisung der Hintergrundfarbe ermöglicht es der Komponente, reaktiv auf Änderungen des value zu reagieren und die Farbe entsprechend anzupassen.

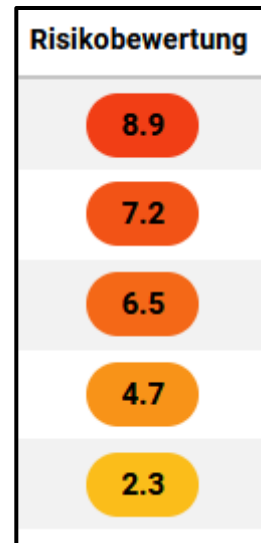


Abbildung 28:  
Risikoindikator

```
computed: {
  color() {
    if (this.value < 1) {
      return "#ffe81c";
    } else if (this.value < 2) {
      return "#fdd31b";
    } else if (this.value < 3) {
      return "#fbbd1a";
    } else if (this.value < 4) {
      return "#faa819";
    } else if (this.value < 5) {
      return "#f89318";
    } else if (this.value < 6) {
      return "#f67e18";
    } else if (this.value < 7) {
      return "#f46817";
    } else if (this.value < 8) {
      return "#f25316";
    } else if (this.value < 9) {
      return "#f13e15";
    } else if (this.value < 10) {
      return "#ef2814";
    } else {
      return "#999999"; // grau; eigentlich das dunkelste rot
                      // aber für die Evaluation Aufgabe 2
                      // geändert
    }
  },
},
```

Abbildung 29: Methode zur Bestimmung des Farbwerts

## Split-Button

### Erklärungen



Abbildung 30:  
Split-Button

Es wird eine dynamische Klassenbindung mittels Vue's Binding Directive für die Buttons verwendet, um die Klassen basierend auf dem hover-Zustand dynamisch zu ändern. Wenn der hover-Zustand true ist, erhalten die Buttons zusätzliche Stile durch die Aktivierung der CSS-Klassen `btn-left-active` und `btn-right-active`.

Die Komponente `SplitButton.vue` implementiert einen geteilten Button mit einer Dropdown-Funktionalität. Wenn der Nutzer über den rechten Button fährt, wird das Dropdown-Menü aktiviert. Beim Klicken auf einen der Links im Dropdown-Menü oder auf den linken Button wird ein `openPopup`-Ereignis mit unterschiedlichen Parametern ausgelöst.

Die Logik, die das Dropdown-Menü steuert, basiert auf dem hover-Zustand mittels eines Boolean. Die Ereignisse `@mouseenter` und `@mouseleave` werden zum Setzen und Zurücksetzen des hover-Zustands verwendet. Die Entscheidung auf diese Variante statt eines Klick-Ereignisses fiel aufgrund der Größe des rechten Teils des Split-Buttons. Die Nutzer müssen der Maus somit nur den Bereich des rechten Teils treffen anstatt den Bereich treffen und Klicken zu müssen. Das soll den Flow und damit die Gebrauchstauglichkeit erhöhen.

Es wird eine dynamische Klassenbindung mittels Vue's Binding

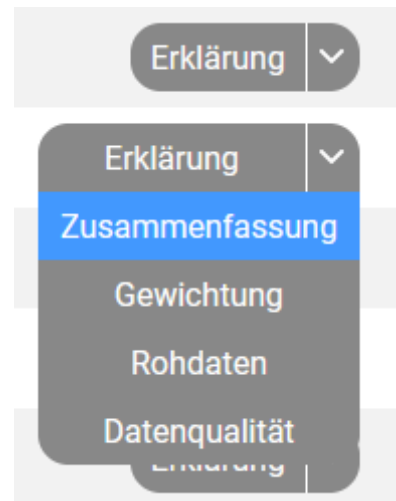


Abbildung 31: Dropdown  
des Split-Buttons

```

<button
  class="btn-left"
  :class="{ 'btn-left-active': hover }"
  @click="$emit('openPopup', 1, contentID)"
>
  Erklärung
</button>

<button
  :class="{ 'btn-right-active': hover }"
  @mouseenter="hover = true"
  @mouseleave="hover = false"
  class="btn-right"
>
  <font-awesome-icon :icon="['fas', 'angle-down']" />
</button>

```

Abbildung 32: Implementierung des Split-Buttons

Die Komponente erhält außerdem eine Zahl als contentID-prop. Zusammen mit der Auswahl aus dem Dropdown wird bei einem Klick die Wahl des Nutzers, welche Ansicht geöffnet werden soll, an die openPopup-Methode übergeben.

```

<a @click="$emit('openPopup', 1, contentID)">Zusammenfassung</a>
<a @click="$emit('openPopup', 2, contentID)">Gewichtung</a>
<a @click="$emit('openPopup', 3, contentID)">Rohdaten</a>
<a @click="$emit('openPopup', 4, contentID)">Datenqualität</a>

```

Abbildung 33: Klick-Events für die Wahl der Erkläransichten

### Gewichtungsanzeige

Die Gewichtungsansicht besteht aus Dropdown rechts und Balkendiagrammen links. Das erste Balkendiagramm zeigt die Summe aller Gewichtungen der einzelnen Datenquellen. Die Balkendiagramme darunter zeigen die Einzelgewichtung der Datenquelle rechts davon. Bei der Implementierung von Diagrammen mit Chart.js sind zwei Elemente zentral: das options- und das data-Prop definieren das Diagramm und werden an Chart.js übergeben.

```

<Bar :options="chartOptions" :data="chartData"/>

```

Abbildung 34: Übergabe Props an Chart.js

Das data-Prop enthält die eigentlichen Daten und bestimmt, welche Daten in der Grafik visualisiert werden. Innerhalb von dem Prop, das als chartData übergeben wird, befindet sich datasets. Dies ist ein Array von Objekten, wobei jedes Objekt eine Gruppe von Datenpunkten repräsentiert, die auf der Grafik dargestellt werden. Jedes dataset-Objekt kann verschiedene Eigenschaften enthalten: die Werte und optische Eigenschaften. Im Falle des summierten Balkendiagramms enthält dataset zusätzliche Logik, um jeden einzelnen Wert mit einer abwechselnden Hintergrundfarbe darzustellen.

```
this.chartData = {
  labels: ["Gesamtwert"],
  datasets: dataFromJson.map((value, index) => ({
    data: [value],
    backgroundColor: index % 2 === 0 ? "#66829C" : "#9FBCCC",
    borderRadius:
      index === 0
        ? [{ topLeft: 30, bottomLeft: 30 }]
        : index === dataFromJson.length - 1
        ? [{ topRight: 30, bottomRight: 30 }]
        : {},
    borderSkipped: false,
    barThickness: 40,
  })),
};
```

Abbildung 35: Konfiguration des summierten Balkendiagramms



Abbildung 36: die zwei Typen Balkendiagramme der Gewichtungsansicht

Für die optische Erscheinung der Balkendiagramme für die Einzelwerte musste ein Trick angewandt werden. Um den Rahmen mit den runden Ecken zu erzeugen, mussten zwei Balkendiagramme übereinandergelegt werden, wobei der zweite Balken leer ist. Dafür wird das Chart.js Plug-In ChartsjsPluginStacked100 verwendet. Das dataset-Array muss in dem Fall mit zwei Items bestückt werden.

```

datasets: [
  {
    data: dataFromJsonProzent, //ein Array
    backgroundColor: index % 2 === 0 ? "#66829C" : "#9FBCCC",
    borderRadius: {
      topLeft: 25,
      bottomLeft: 25,
    },
    borderSkipped: false,
    barThickness: 40,
    borderWidth: 0,
  },
  {
    data: [100 - dataFromJsonProzent[0]],
    backgroundColor: "transparent",
    borderWidth: 1,
    barThickness: 40,
    borderRadius: {
      topLeft: 25,
      topRight: 25,
      bottomLeft: 25,
      bottomRight: 25,
    },
  },
],

```

Abbildung 37: dataset Konfiguration für stacked bar charts

Um sicherzustellen, dass für den leeren Balken keine Labels und Tooltips angezeigt werden, muss für das options-Prop Logik implementiert werden:

```

formatter: (value, context) => {
  // Prüfen, ob das aktuelle Dataset das erste ist
  if (context.datasetIndex === 0) {
    return `${value}%`; // Nur für das erste Dataset das Label anzeigen
  } else {
    return null; // Kein Label für andere Datasets
  }
},

```

Abbildung 38: Konfiguration der speziellen Labelanzeige in Chart.js

## 5.5 Fazit

In der Implementierung konnte ein High-Fidelity Prototyp ausgehend von der zweiten Iteration der Konzeption umgesetzt werden. Aus Zeitgründen wurde keine Implementierung der Kartenansicht umgesetzt. Dies hatte laut Anforderungsliste auch keine hohe Priorität (siehe 4.3). Außerdem wurde keine Logik für die Eingabe des Datums und Ausgabe der Listeneinträge implementiert, da sich im Laufe der Realisierung herausstellte, dass dies für Evaluierung des Systems nicht weiter benötigt wurde. Diese und weitere Aspekte werden in den Abschnitten Offene Punkte (8.1) und Ausblick (8.2) näher erläutert.

## 6 Dialogbeispiele

Dieses Kapitel stellt eine typische Bedienung des Systems dar, wie sie auch im Rahmen der Evaluation durch die Nutzer erfolgen konnte. Die Reihenfolge ist chronologisch.

Zu Beginn befindet sich der Nutzer auf der Hauptansicht. Ihm wird eine Liste mit zu kontrollierenden Betrieben angezeigt. Die Liste enthält neben verschiedenen statischen Informationen zum Betrieb auch die KI-generierte Risikobewertung.

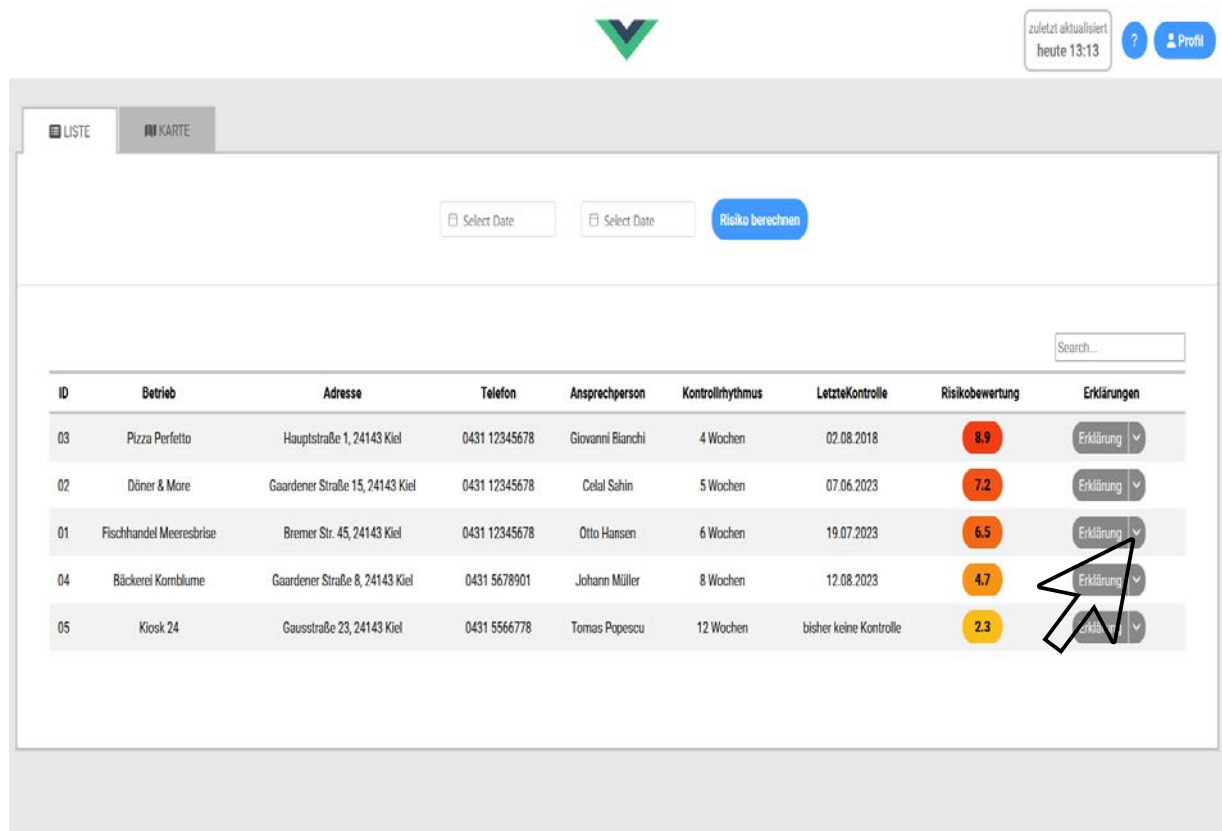


Abbildung 39: Implementierte Listenansicht

Der Nutzer möchte beispielsweise wissen, wie die einzelnen Datenquellen gewichtet und bewertet werden. Er öffnet den Dropdown und klickt den Button „Gewichtung“.

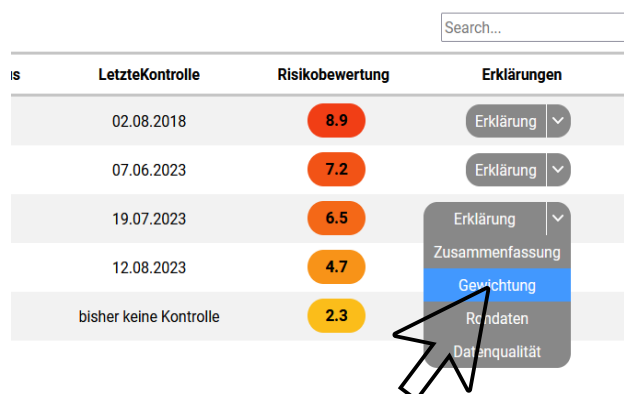


Abbildung 40: Implementiertes Dropdown-Menü

Es öffnet sich die Ansicht für die Gewichtung. Dem Nutzer wird eine Übersicht der für die Bewertung genutzten Datenquellen (rechts) und die Gewichtungsanzeige (links) mit den Prozents angezeigt. Der Nutzer möchte nun weiterhin wissen, wie eine spezifische Datenquelle bewertet wurde. Dafür klickt er auf den Titel der spezifischen Quelle. Es öffnet sich der Dropdown.



Abbildung 41: Implementierte Gewichtungsansicht

Im Dropdown wird neben den unterschiedlichen Informationen aus und zu der Datenquelle auch die Risikobewertung dieser spezifischen Datenquelle angezeigt.

Wenn dem Nutzer das noch nicht reicht, kann er von hier aus direkt zu den Rohdaten dieser Datenquelle, also zu den Daten, die als Input für das KI-System dienen. Dafür klickt der Nutzer unten auf den Button „zur Datenquelle“.

Kontrollbericht vom 22.03.2022

Die Bewertung der Datenquelle "Kontrollbericht vom 22.03.2022" basiert auf der Auswertung der KI.

Die Risikobewertung für die spezifische Datenquelle "Kontrollbericht vom 22.03.2022" ergibt einen Wert von **9.1**

Allgemeine Gewichtung der Datenquelle: sehr hoch

Kontrollzweck: planmäßige Routinekontrolle am 22.03.2022.

Identifizierte Themen und Informationen:

- Verkaufsraum & Kühltheke:
- Technischer Defekt in der Kühlkette: Unterbrechung der Kühlkette im Verkaufsraum für unbekannte Dauer
- Anordnung der Reparatur der Kühlgeräte und entsorgung der betroffenen Lebensmittel.
- Anzeichen für Feuchtigkeit und Schimmel in einigen Bereichen der Kühltheke
- Umgehende Reinigung und Desinfektion der gesamten Kühltheke wurde angeordnet
- Mangelhafte Umsetzung des HACCP-Konzepts: Fehlende nachweisbare Dokumentationen kritischer Kontrollpunkte.
- Anordnung zur Implementierung eines vollständigen HACCP-Konzepts

zur Datenquelle



Abb. 42: Implementierte Einzelbewertung Dropdown



Die Ansicht wechselt zur Rohdatenansicht „Alle Daten“. Der spezifische Dropdown, in dem sich die gesuchten Informationen befinden, wurde beim Wechsel bereits geöffnet.

Damit der Nutzer sichergehen kann, dass er dabei ist die richtige Datenquelle zu öffnen oder falls er nur eine zentrale Information aus der Quelle benötigt, befindet sich im Dropdown auch eine kurze Zusammenfassung des Inhalts.

## Alle Daten



Betrieb Fischhandel Meeresbrise - Risikobewertung 6.5

### Datenquelle Betriebskontrolle 1

Das ist eine vom System erstellte Zusammenfassung der Datenquelle. Die Originalquelle befindet sich am Ende dieser Zusammenfassung.

#### Kontrollbericht Nr. 1

**Betrieb:** Fischhandel Meeresbrise

**Kontrolldatum:** 22.03.2022

**Kontrollobjekt:** ID-9834-KL

**Kontrollzweck:** planmäßige Routinekontrolle am: 22.03.2022

**Kontrollierte Räume/Bereiche:** Betriebsstätte (allgemein), Küche, Verkaufsraum

**Nicht kontrollierte Räume/ Bereiche:** Personalumkleide, Kühllager

**HACCP-Konzept:** nicht vorhanden

#### Feststellungen:

**Verkaufsraum:**

- **Feststellung:** Die Kühlkette war aufgrund eines technischen Defekts im Verkaufsraum für eine unbekannte Dauer unterbrochen. Die Temperatur der kühlbedürftigen Produkte nicht durchgehend den gesetzlichen Vorschriften entsprach.

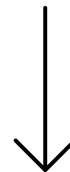


Abbildung 44: Implementierte Rohdatenansicht

Der Nutzer möchte aber in die Rohdaten für diese Datenquelle reinschauen und scrollt dafür runter ans Ende des Dropdowns.

Dort findet er eine Miniaturansicht des Scans des Kontrollprotokolls. Er klickt auf die Ansicht, um sie zu vergrößern.

- **Feststellung:** Anzeichen von Feuchtigkeit und S
  - Anordnung:** Umgehende Reinigung und Desinf
  - **Feststellung:** Das HACCP-Konzept des Betriebs Risiko für die Lebensmittelsicherheit darstellt.
  - Anordnung:** Die Implementierung eines vollstän
- Weitere Informationen:
- Es wurden Lichtbilder wurden angefertigt, Merk



Abbildung 43: Implementierte Miniaturansicht

Die Miniaturansicht hat sich vergrößert und ist nun für den Nutzer einseh- und scrollbar.



Abbildung 45: Schließen der vergrößerten Miniaturansicht

Wenn der Nutzer fertig ist, kann er durch einen Klick auf den blauen Close-Button sowohl in der Vergrößerungsansicht als auch in der übergeordneten Ansicht zurück zur Listenansicht.

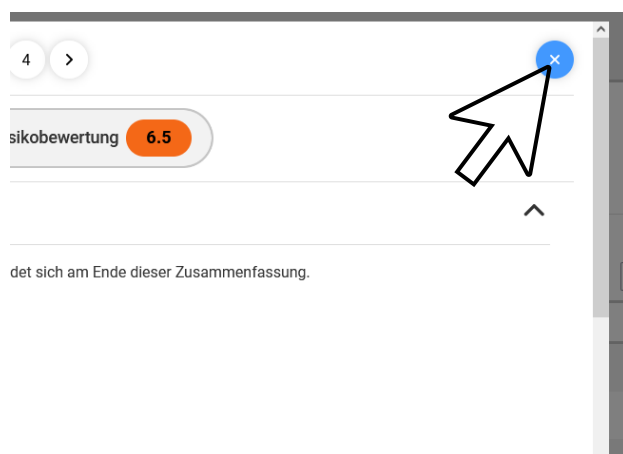


Abbildung 46: Schließen des Popups

## 7 Summative Evaluation

Die summative Evaluation repräsentiert die abschließende Phase im menschenzentrierten Gestaltungsprozess. Sie soll feststellen, ob das entwickelte System den vorgegebenen Anforderungen entspricht. Im Kapitel 7.1 werden die Ziele der Evaluation und die zentrale Fragestellung vorgestellt. Kapitel 7.2 beschreibt die Methodik der Evaluation mittels A/B-Test und ausgewählten Fragebögen. In Kapitel 7.3 werden die erhobenen Daten als Ergebnisse präsentiert und anschließend in Kapitel 7.4 diskutiert. Abschließend folgt in Kapitel 7.5 ein Fazit der durchgeführten Evaluation.

### 7.1 Ziel

Die Ziele der summativen Evaluation ergeben sich aus den Zielen dieser Bachelorarbeit (Kapitel 1.1). Das primäre Evaluationsziel ist die Überprüfung, ob der entwickelte Prototyp tatsächlich zu einer höheren Nachvollziehbarkeit der KI-Ergebnisse bei seinen Nutzern beiträgt. Die Nachvollziehbarkeit kann dabei nach Schrills & Franke (2022) anhand der folgenden Kategorien operationalisiert werden:

1. **Transparenz über Ergebnisse des KI-Systems:** Es wird untersucht, inwieweit das Interface klare und verständliche Informationen über die von dem KI-System erzeugten Ergebnisse bereitstellt.
2. **Verständnis des Ergebnisses des KI-Systems:** Es wird ermittelt, ob die Nutzer in der Lage sind, die Arbeitsweise des KI-Systems und die zugrunde liegenden Entscheidungsprozesse zu verstehen.
3. **Vorhersehbarkeit des Ergebnisses:** Es wird analysiert, ob die Nutzer in der Lage sind, die Ergebnisse des KI-Systems in gewissem Maße vorherzusagen, was auf eine hohe Nachvollziehbarkeit hinweisen würde.

Ein weiteres Ziel der Evaluation besteht darin, die Gebrauchstauglichkeit des Prototyps zu messen. Dies ist entscheidend, da eine hohe Nachvollziehbarkeit allein nicht ausreicht, wenn die Benutzer die Anwendung nicht effektiv und zufriedenstellend nutzen können. Darüber hinaus kann eine nicht ausreichende Usability die Ergebnisse der Evaluierung der Nachvollziehbarkeit verzerren.

Zudem soll die Nutzerbeanspruchung durch das System gemessen werden, denn das implementierte System soll nicht nur die Transparenz und Verständlichkeit des fiktiven KI-Systems erhöhen, sondern auch eine geringe Belastung für die Nutzer während der Verwendung darstellen.

Aus den Zielen lässt sich folgende Fragestellung herleiten:

Inwiefern beeinflusst ein System, das auf Anforderungen für ein nachvollziehbares KI-System (XAI) basiert, im Vergleich zu einem System ohne diese Grundlage, die Nachvollziehbarkeit?

Diese Fragestellung zielt darauf ab, die Effektivität des XAI-basierten Systems hinsichtlich der Nachvollziehbarkeit zu bewerten. Die Vermutung besteht darin, dass ein nach XAI-Anforderungen gestaltetes System die Nachvollziehbarkeit positiv beeinflusst.

## 7.2 Methode

Im nächsten Abschnitt werden die verschiedenen Aspekte der Methodik vorgestellt und erläutert.

### 7.2.1 Design

Das Design der in dieser Arbeit umgesetzten Evaluation folgt dem Ansatz einer multi-methodischen Evaluationsstudie. Da die Parameter einer A-priori-Teststärkenanalyse schwer abzuschätzen sind und die Akquise der vermutlich benötigten Anzahl von Testpersonen für diese Arbeit nicht umsetzbar ist, kann das Design explizit nicht als Experiment konzipiert werden. Somit kann die Evaluation auch nicht für sich beanspruchen kausale Zusammenhänge nachzuweisen, sondern vielmehr korrelative Tendenzen und Nutzerwahrnehmungen aufzuzeigen, die für weiterführende Studien als Ausgangspunkt dienen können.

Der Schwerpunkt liegt dabei auf der Beurteilung zweier zentraler Kriterien: der Nachvollziehbarkeit und der Usability des zu testenden Systems. Diese Kriterien werden primär durch Fragebögen erfasst (siehe 7.2.3. „verwendete Fragebögen“). Sie werden außerdem im selben Durchgang einer Evaluationseinheit evaluiert. Die Kombination beider Elemente schont die zeitlichen Ressourcen der Teilnehmer.

Basierend auf den Empfehlungen von Lazar et al. (2017) wird ein experimentelles A/B-Test-Design implementiert, bei dem zwei Systemvarianten gegeneinander getestet werden. Die aus der Operationalisierung abgeleiteten Metriken werden für beide Systeme erhoben und verglichen. Zugänglich werden die Metriken durch die Bearbeitung verschiedener Aufgaben durch die Teilnehmer (siehe 7.2.4). Diese Vorgehensweise ermöglicht es weitere Daten zu erfassen, wie beispielsweise die Zeitmessung der Dauer der Aufgabenbewältigung und die Nutzerbeanspruchung durch die jeweiligen Aufgabe mittels Fragebogen (siehe 7.2.3. „verwendete Fragebögen“).

Ergänzend wird während der Evaluation die Think-Aloud-Methode (Nielsen, 1994) angewandt und so zusätzlich qualitative Daten generiert. Sie dient dazu, tiefere Einblicke in das Denken der

Teilnehmer zu erhalten und kann aufzeigen, worin bestimmte Probleme (für den spezifischen Teilnehmer) begründet liegen.

Aufgrund eines Mangels an Teilnehmern, wird ein Within-Subjects-Design (Lazar et al., 2017) gewählt. Dies bedeutet, dass dieselben Teilnehmer beide Bedingungen der Studie durchlaufen. Dies hat den Vorteil, dass weniger Teilnehmer benötigt werden. Außerdem dient jede Versuchsperson als ihre eigene Kontrollinstanz, was die Auswirkungen von störenden Variablen, die zwischen den Individuen variieren könnten, wie Vorerfahrungen oder Motivation, minimiert. Nachteilig am Within-Subjects-Design ist das mögliche Auftreten von Reihenfolgeeffekten, bei dem die Reihenfolge der Bedingungen die Ergebnisse beeinflusst. Die Aufgabenbewältigung zunächst ohne und anschließend mit der Unterstützung durch XAI könnte zu einer Verbesserung der Leistung in der zweiten Runde führen. Diese Verbesserung könnte jedoch primär auf die Wiederholung zurückzuführen sein, nicht unbedingt auf den Einfluss von XAI. Eine Umkehrung der Reihenfolge könnte einen Lerneffekt zur Folge haben. Die Teilnehmer könnten aus der ersten Runde mit XAI lernen und ihre Leistung in der zweiten Runde verbessern. Um diese Probleme zu vermeiden, wird die Reihenfolge der Bedingungen randomisiert, also zufällig bestimmt, wer zuerst welche Variante nutzt. Dies trägt dazu bei Lerneffekte und Reihenfolgeeffekte zu minimieren, da diese bei der Berechnung der Mittelwerte gleichmäßig über die Gruppen verteilt wären (Nielsen, 1994).

## 7.2.2 Teilnehmer

Die Gewinnung von Teilnehmern für die Evaluation stellte eine Herausforderung dar. Dies machte die Beantwortung der Frage nach der Mindestanzahl der Teilnehmer umso wichtiger. Nielsen (2000) schlägt für die meisten Usability-Tests eine Anzahl von fünf Teilnehmern vor.

Seine Forschungen zeigen, dass eine Gruppe in dieser Größe in der Regel etwa 85% der Probleme eines Designs identifizieren kann. Er unterstreicht, dass die signifikanten Usability-Probleme tendenziell unabhängig von der Nutzerzahl auftauchen und dass zusätzliche Erkenntnisse ab dem fünften Nutzer zunehmend rarer werden.

Es muss jedoch angemerkt werden, dass Niensens Empfehlung zum einen einschließt, dass mehrere Iterationen des Tests durchgeführt werden und zum anderen die Empfehlungen für qualitative Usability-Tests gilt, deren Hauptziel die Identifizierung von Problemen ist, anstatt statistisch signifikante Resultate zu liefern. Für quantitative Studien, wie A/B-Tests, die auf statistische Generalisierbarkeit abzielen, ist in der Regel eine größere Anzahl von Teilnehmern erforderlich.

Aufgrund der vermuteten Reihenfolgeeffekte bzw. Lerneffekte (siehe 7.2.1 Design), musste die Reihenfolge der Systemnutzung im A/B Test randomisiert sein, um die Effekte zu minimieren. Das hat es ermöglicht beide Gruppen – beginnt mit XAI oder ohne – aktiv miteinander zu vergleichen.

Daraus folgte, dass eine Verdopplung der Teilnehmeranzahl nötig wäre, was zu einer angestrebten Mindestanzahl von zehn Teilnehmern führte.

Außerdem muss zu der Auswahl der Teilnehmer gesagt werden, dass die demographische Zusammensetzung nicht so heterogen ist, wie es idealerweise der Fall sein sollte, und anzunehmen ist, dass sie sich von der Zusammensetzung der eigentlichen Zielgruppe unterscheidet. Das durchschnittliche Alter der Teilnehmenden ist durchschnittlich 30,4 Jahre (SD=5,7), überwiegend männlich und IT-erfahren. 3 von 10 Personen gaben an keine Vorerfahrung mit KI-Systemen zu haben (alle weiblich).

Die folgende Tabelle listet die Teilnehmer und ihre Daten auf:

<b>TN-ID</b>	<b>Alter</b>	<b>Geschlecht</b>	<b>Beruf</b>	<b>Vorerfahrung mit KI- Systemen</b>	<b>Begann mit System</b>
<b>1</b>	27	männlich	Masterstudent Medieninformatik	Ja	XAI
<b>2</b>	37	männlich	Programmierer	Ja	Nicht-XAI
<b>3</b>	33	weiblich	Richterin	Nein	XAI
<b>4</b>	27	männlich	Student	Ja	Nicht-XAI
<b>5</b>	25	männlich	Medieninformatiker	Ja	Nicht-XAI
<b>6</b>	32	weiblich	Gewerkschafts- sekretärin	Nein	XAI
<b>7</b>	28	weiblich	Gesundheits- & Krankenpflegerin	Nein	Nicht-XAI
<b>8</b>	20	männlich	Student Wirtschafts- informatik	Ja	XAI
<b>9</b>	38	männlich	Designer	Ja	XAI
<b>10</b>	37	männlich	DevOps-Engineer	Ja	Nicht-XAI

*Tabelle 3: Liste der Evaluationsteilnehmer*

### 7.2.3 Setting und Instrumente

#### **Setting**

Die Evaluation des Systems zeichnete sich durch eine dezentrale Durchführung aus, die sich aus der geografischen Verteilung der Teilnehmenden über verschiedene Städte ergab. Das brachte es

mit sich, dass die Evaluationsdurchgänge in unterschiedlichen Umgebungen stattfanden. Da verschiedene Fragebögen zu unterschiedlichen Momenten während der Evaluation ausgefüllt werden sollten und auch sichergestellt werden musste, dass die Teilnehmer ihre Rolle und die Aufgaben verstanden haben, war der Ersteller als Versuchsleiter während der Durchführung anwesend. Der Großteil der Evaluationen wurde online in privater Umgebung durchgeführt und über Screensharing mitverfolgt. Ein paar Durchgänge fanden zu Hause beim Ersteller statt.

### **Verwendete Fragebögen**

Dieser Abschnitt stellt die verwendeten Fragebögen vor. Die vollständigen Fragebögen sind im Anhang zu finden.

#### *Subjective Information Processing Awareness-Skala*

Die subjektive Informationsverarbeitungswahrnehmung (SIPA) beschreibt die Erfahrungen der Benutzer während ihrer Interaktion mit intelligenten Systemen und basiert auf der Theorie des Situationsbewusstseins nach Endsley von 1988 (zitiert nach Schrills et al., 2021). SIPA kann definiert werden als die Erfahrung, durch ein System in die Lage versetzt zu werden, dessen Informationsverarbeitung wahrzunehmen, zu verstehen und seine Informationsverarbeitung vorherzusagen. Die SIPA-Skala ist laut den Erstellern eine ökonomische, eindimensionale Skala, die sich für eine Anwendung in der Explainable Artificial Intelligence eignet (Schrills et al., 2021). Die SIPA-Skala kann in Kontexten angewendet werden, in denen Menschen mit intelligenten Systemen zusammenarbeiten. Sie ist als Indikator für Systemdesigns gedacht, die explizite Erklärungen enthalten können oder auch keine Erklärungen enthalten. Die Ergebnisse der SIPA-Skala sollten nicht ohne Bezug (d.h. eine Kontrollbedingung oder bestimmte Zeitpunkte) interpretiert werden.

#### *System Usability Scale -Fragebogen*

Neben dem SIPA-Fragebogen wurde außerdem die System Usability Scale (SUS) (Brooke, 1995) genutzt, um die Gebrauchstauglichkeit des Systems zu evaluieren. SUS ist eine etablierte Skala, die zur Evaluierung der Gebrauchstauglichkeit verschiedener Systeme eingesetzt werden kann. In dieser Untersuchung dient der SUS dazu, zu überprüfen, ob die Nachvollziehbarkeit des implementierten Systems durch Usability-Probleme beeinträchtigt wird, anstatt durch eine unzureichende Umsetzung der XAI-Anforderungen, bzw. um sicherzustellen, dass eventuelle Schwächen in der XAI-Konzeption nicht fälschlicherweise auf Mängel in der Gebrauchstauglichkeit zurückgeführt werden.

Die Skala besteht aus 10 Items, die mittels einer 5-Punkt Likert-Skala zur Einschätzung der Zustimmung beurteilt werden. Die Ergebnisse werden dabei so codiert und verrechnet, dass das Ergebnis einen Wert zwischen 0 (niedrigste Gebrauchstauglichkeit) und 100 (höchste Gebrauchstauglichkeit) ergibt. Für den Fragebogen wurde eine bereits vorhandene deutsche Übersetzung (Rummel, 2013) der Skala verwendet. Die Ergebnisse des SUS geben einen schnellen Überblick darüber, wie Benutzer ein System wahrnehmen, und helfen bei der Identifizierung von Bereichen, die verbessert werden können.

### *NASA Task Load Index Fragebogen*

Der NASA-TLX Fragebogen ist ein etabliertes Bewertungsinstrument, das entwickelt wurde, um die wahrgenommene Arbeitsbelastung von Personen bei der Ausführung verschiedener Aufgaben zu messen (NASA TLX, 2022). Entwickelt von der Human Performance Group der NASA, bietet der NASA-TLX eine multidimensionale Skala, die sechs verschiedene Aspekte der Arbeitsbelastung bewertet: Geistige Anforderungen, körperliche Anforderungen, zeitliche Anforderungen, Leistung, Anstrengung und Frustration. Die Teilnehmer bewerten diese Aspekte basierend auf ihrer persönlichen Erfahrung während einer spezifischen Aufgabe. Ein hoher Wert kann Aufschluss darüber geben, ob das System oder der erstellte Inhalt falsch konzipiert wurden.

### **Verwendete Anwendung**

Für den A/B-Test wurden zwei verschiedene Versionen der entwickelten Anwendung genutzt: Das für diese Arbeit entwickelte XAI-System und eine modifizierte Variante dieser Anwendung, aus der gezielt die XAI-Funktionen bzw. Eigenschaften entfernt wurden, um einen Vergleich zu ermöglichen.

Die Version ohne XAI-Features wurde auf zwei wesentliche Ansichten reduziert. Zum einen gab es die Listenansicht, die in ihrer Grundstruktur dem XAI-System ähnelte, da die farblichen Risiko-Indikatoren erhalten blieben. Obwohl das aktuell in der Lebensmittelüberwachung verwendete System Balvi IP solche visuellen Hilfsmittel nicht standardmäßig anbietet, wurden sie drin gelassen, um die Teilnehmer nicht zu überfordern, da der Kontext für sie völlig neu ist. Zum anderen bestand die zweite Ansicht aus einer Darstellung der Rohdaten, die in dieser Version bewusst auf eine „unverarbeitete“ Form beschränkt wurden. Während im XAI-System die „Rohdaten“ für eine bessere Nachvollziehbarkeit durch das KI-System optimiert sind und die eigentlichen „Rohdaten“ aus einer Miniaturansicht heraus in einer Vollansicht dargestellt werden können, wurden in der Nicht-XAI-Variante die optimierten Daten durch die „Rohdaten“ aus der Miniaturansicht getauscht und die Miniaturansicht damit entfernt. Dies sollte die Unterschiede im Nutzererlebnis zwischen den beiden Systemen hervorheben und analysierbar machen.

### **Verwendete Datensätze**

Da XAI-Systeme Erklärungen und Begründungen für die Anwender bereitstellen, die die KI-Prozesse und Ergebnisse erklären, mussten Inhalte produziert werden, die im Kontext der Lebensmittelüberwachung sinnvoll sind. Die Vorgehensweise war iterativ mit mehreren Durchgängen.

Die Herausforderung bei der Gestaltung der Datensätze war es die Daten sinnvoll auf mehreren Ebenen erscheinen zu lassen: Sie sollten den Anschein von KI-generierten Daten erwecken, im Kontext der Lebensmittelüberwachung für die Versuchsteilnehmer glaubwürdig sein und über alle erstellten Datenquellen hinweg konsistent auf den spezifischen Betrieb abgestimmt sein.

Als Grundlage dienten Scans von echten Kontrollprotokollen, die von Lebensmittelkontrolleuren erstellt und vom Betreuer dieser Arbeit bereitgestellt wurden. Da diese Daten nicht im Original verwendet werden durften, wurde eine eigene Version für diese Arbeit erstellt. Diese Version



beinhaltet im Wesentlichen alle relevanten Aspekte der Originalprotokolle und repräsentiert die Kontrollprotokolle im entwickelten System.

Wie die Analyse ergeben hat, sind neben Kontrollprotokollen auch Gewerbeanmeldedaten sehr relevant für Lebensmittelkontrolleure, sodass für jeden fiktiven Betrieb auch diese Daten erstellt wurden.

Um die Konsistenz zu erzeugen, wurde für jeden Betrieb eine Art Historie bzw. Verlauf der hygienischen Zustände erstellt. Dies berücksichtigte, dass sich die hygienischen Zustände der Betriebe im Laufe der Zeit verbessern oder verschlechtern können, was sich in den Kontrollergebnissen widerspiegelt.

Der Teil der Informationen zu einem Betrieb, der in der Zukunft als zusätzliche Datenquelle für die Bewertung relevant sein könnte, sind die Onlinerezensionen, die von dem KI-System aus dem Internet gesammelt werden. Dafür wurden solche Rezensionen online recherchiert und nach diesem Vorbild erstellt.

Ein typischer Datensatz für einen Betrieb besteht somit aus:

- 1 bis 3 Kontrollprotokollen
- 1 Gewerbeanmeldedaten
- Ein paar Onlinerezensionen /-beschwerden

Zusätzlich wurde für ein Betrieb eine „Beschwerde aus der Bevölkerung“ erstellt.

Nach dem diese Daten vollständig waren, wurden die „KI-generierten“ Zusammenfassungen geschrieben. Hierfür wurden wichtige Informationen aus den verschiedenen Datenquellen zusammengesammelt und in einer möglichst neutralen und technischen Ausdrucksweise zusammengefasst. In einer ersten Version enthielt eine solche Zusammenfassung ca. 350 Zeichen. Ein Vortest zur Evaluation ergab jedoch, dass dies als zu umfangreich empfunden wurde von der Testperson, woraufhin die durchschnittliche Zeichenlänge auf ungefähr 250 Zeichen verringert wurde.

Um den Nutzern des Systems nicht nur die hygienischen Zustände in Betrieben aus Sicht des KI-Systems zur Verfügung zu stellen, sondern auch Einsichten in den Bewertungsprozess des Systems, wurde für die Gewichtungsansicht Gewichtungsdaten in Form von Arrays mit Prozentzahlen erstellt. Jede Datenquelle erhielt einen spezifischen Gewichtungswert. Diese variieren für jeden Betrieb, abhängig davon, welche Datenquellen vorhanden sind und in welchem Umfang sie vorliegen.

Um das KI-System interessanter zu gestalten und das Potenzial zukünftiger Entwicklungen aufzuzeigen, wurden Daten integriert, die Zusammenhänge aufdecken, die Menschen nicht unmittelbar erkennen können. Das sind beispielsweise statistische Zusammenhänge, die nicht unbedingt aus den einzelnen Rohdaten ersichtlich sind. Diese Informationen befinden sich in den Übersicht-Dropdowns in den Gewichtungsansichten und in den Zusammenfassungen. Außerdem

wurden kurze Textinformationen zu den Ampelfarben der Datenqualität verfasst, die Details zu der farblichen Bewertung der spezifischen Kategorie liefern. Ein Beispiel Datensatz, der von jeder Art Datenquelle eins enthält, befindet sich im Anhang. Die gesamten Datensätze sind im System einsehbar.

## 7.2.4 Ablauf

Im folgenden Abschnitt werden die verwendeten Aufgaben und der Ablauf der Evaluation beschrieben.

Um die Metriken erfassen zu können wurden Aufgaben konzipiert, die die Teilnehmer in die Lage versetzen sollen, sowohl das System bewerten zu können als auch eine Selbstbewertung ihrer Wahrnehmung vorzunehmen. Die Aufgaben orientieren sich an der Operationalisierung durch Schrills (2021) (Kapitel 7.1), wobei zu beachten ist, dass nicht jede Kategorie der Operationalisierung durch eine spezifische Aufgabe abgedeckt wurde. Beispielsweise wurde die Kategorie der „Klarheit über die Ergebnisse des KI-Systems“ allein durch den SIPA-Fragebogen erfasst. Bei den beiden anderen Kategorien werden sowohl die Fragen aus dem SIPA-Fragebogen als auch zusätzliche Metriken erfasst.

Die erste Aufgabe zielt darauf ab, das Verständnis der Teilnehmer für die Berechnung der Ergebnisse des KI-Systems zu erfassen. Ihnen wird die Aufgabe gestellt, die Informationen zu den gegebenen Betrieben nachzuvollziehen und daraufhin ein Multiple-Choice-Quiz zu beantworten, das Fragen zur Funktionsweise des KI-Systems umfasst.

In der zweiten Aufgabe geht es um das Erkennen von Fehlern im Modell. Auch diese Aufgabe zielt, wie Aufgabe 1, auf das Verständnis der Teilnehmer ab. In beiden Systemvarianten wurden die Datensätze für jeweils eines Betriebs teilweise verfälscht. Die Teilnehmer sollten beurteilen, ob alle Priorisierungen und Berechnungen des KI-Systems korrekt und sinnvoll sind oder ob für bestimmte Betriebe unangemessene Werte berechnet wurden. Während das System mit XAI den Teilnehmern beispielsweise über die Ansicht für die Datenqualität Hinweise liefert und damit über die Verlässlichkeit des berechneten Risikowerts informiert, mussten die Teilnehmer mit dem Nicht-XAI-System allein auf der Basis von den „Rohdaten“ entscheiden. Es wird nicht nur festgehalten, ob der Betrieb korrekt identifiziert wurde, sondern auch, wie viel Zeit die Teilnehmer für die Aufgabe benötigten und wie anstrengend sie diese empfanden, wobei letzteres über den NASA TLX-Fragebogen ermittelt wurde.

Die dritte Aufgabe beschäftigt sich mit der Vorhersehbarkeit von KI-Ergebnissen. Die Teilnehmer bekamen die Aufgabe einen Betrieb, der noch nicht priorisiert war, entsprechend ihren Erwartungen an die Funktionsweise des Systems selbst zwischen die bereits priorisierten Betriebe einzuordnen. Auch hier werden das XAI-System und das Nicht-XAI-System verglichen, um zu ermitteln, ob die Verfügbarkeit von mehr Informationen zu Zeitgewinn aufgrund

erleichterter Aufgabenbewältigung führt oder ob sie zu Zeitverlust durch eine Informationsüberflutung beiträgt.

Zusätzlich wird die wahrgenommene Nachvollziehbarkeit mit dem SIPA-Fragebogen und die Gebrauchstauglichkeit mit dem SUS-Fragebogen erfasst. So werden neben den objektiven Ergebnissen der Aufgabenerfüllung auch subjektive Daten erhoben.

### Übersicht des zeitlichen Ablaufs

1. Begrüßung
2. Erläuterung der Evaluation *(insgesamt 5 Minuten)*
  - a. Kurz das Thema vorstellen
  - b. Kurz das Ziel vorstellen
  - c. Vorstellung beider Systeme
3. Einverständniserklärung einholen
4. Erste System: *(insgesamt 50 Minuten)*
  - a. Aufgabe 1 & 2 mit Think Aloud *(max. 15 Minuten)*
    - i. TXL-Fragebogen *(5 Minuten)*
    - ii. Quiz *(5 Minuten)*
  - b. Aufgabe 3 mit Think Aloud *(max. 10 Minuten)*
    - i. TXL-Fragebogen *(5 Minuten)*
  - c. SIPA-Fragebogen *(5 Minuten)*
  - d. SUS-Fragebogen *(5 Minuten)*
5. Zweites System: *(insgesamt 45 Minuten)*
  - a. Aufgabe 1 & 2 mit Think Aloud *(max. 15 Minuten)*
    - i. TXL-Fragebogen *(5 Minuten)*
  - b. Aufgabe 3 mit Think Aloud *(max. 10 Minuten)*
    - i. TXL-Fragebogen *(5 Minuten)*
  - c. SIPA-Fragebogen *(5 Minuten)*
  - d. TXL-Fragebogen *(5 Minuten)*
6. Abschluss
  - a. Gibt es noch Anmerkungen oder Fragen?
  - b. Bedanken und Verabschiedung

## 7.3 Ergebnisse

Im Folgenden werden die Ergebnisse aus der Evaluation detailliert dargestellt.

### 7.3.1 Wahrgenommene Nachvollziehbarkeit (SIPA-Skala)

Die SIPA-Skala wurde verwendet, um die zentrale Fragestellung der summativen Evaluation beantworten zu können. Diese Skala diente dazu, den Grad der Nachvollziehbarkeit zu messen. Die daraus resultierenden Ergebnisse werden nachfolgend dargestellt.

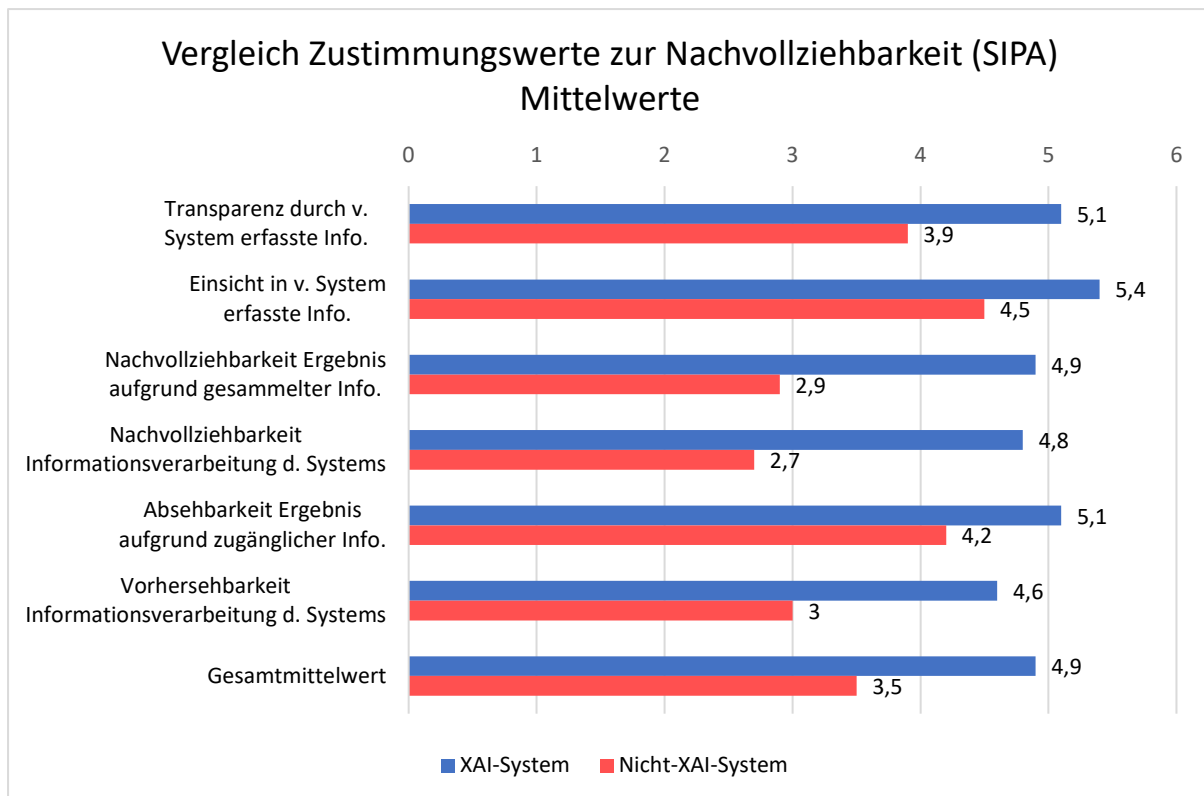
Für die von den Teilnehmern wahrgenommene Nachvollziehbarkeit ergaben sich folgende Mittelwerte und Standardabweichungen:

<b>Gesamtmittelwerte SIPA</b>			
	<b>M</b>	<b>SD</b>	<b>N</b>
<b>XAI-System</b>	4.98	0.44	10
<b>Nicht-XAI-System</b>	3.53	0.65	10

*Tabelle 4: Vergleich Gesamtmittelwerte für die Erhebung der Nachvollziehbarkeit*

Die Mittelwerte repräsentieren den Zustimmungsgrad über alle Aussagen und Teilnehmer. Die Ergebnisse basieren auf einer 6-Punkt-Likert-Skala, wobei 1 „stimme gar nicht zu“ und 6 „stimme völlig zu“ bedeutet.

Abbildung 39 stellt die Ergebnisse beider Systeme zu den einzelnen Aussagen gegenüber:



*Abbildung 47: SIPA Ergebnisse beider Systeme, N=10*

Die zwei nachfolgenden Tabellen zeigen die Ergebnisse bezüglich einzelner Aussagen. Zuerst für das XAI-System und darauffolgend für das Nicht-XAI-System:

#### Ergebnisse für das XAI-System

Kategorie	Aussage	M	SD	N
Transparenz über Ergebnisse des KI-Systems	Es war für mich transparent, welche Informationen vom System erfasst wurden	5.1	0.3	10
	Die Informationen, die das System erfassen konnte, waren für mich einsehbar	5.4	0.66	10
Verständnis des Ergebnisses des KI-Systems	Es war für mich nachvollziehbar, wie die gesammelten Informationen zu diesem Ergebnis führten	4.9	53	10
	Die Informationsverarbeitung des Systems war für mich nachvollziehbar	4.8	1.08	10

Vorhersehbarkeit des Ergebnisses	Mit den für mich zugänglichen Informationen war das Ergebnis für mich absehbar	5.1	0.3	10
	Die Informationsverarbeitung des Systems war für mich vorhersehbar	4.6	0.8	10

*Tabelle 5: Ergebnisse Einzelaussagen zur Nachvollziehbarkeit (XAI-System)*

Sowohl der Mittelwert über alle Aussagen als auch die Mittelwerte der Einzelaussagen liegen ungefähr im oberen Drittel der Bewertungsskala.

Ergebnisse für das Nicht-XAI-System				
Kategorie	Aussage	M	SD	N
Transparenz über Ergebnisse des KI-Systems	1	3.9	1.3	10
	2	4.5	1.1	10
Verständnis des Ergebnisses des KI-Systems	3	2.9	0.9	10
	4	2.7	1.2	10
Vorhersehbarkeit des Ergebnisses	5	4.2	0.6	10
	6	3.0	1.1	10

*Tabelle 6: Ergebnisse Einzelaussagen zur Nachvollziehbarkeit (Nicht-XAI-System)*

### 7.3.2 Beanspruchung durch Bearbeitung der Aufgaben (Task Load Index)

#### Aufgabe 2: „Modellfehler finden“

Die Gesamtbeanspruchung wurde anhand der Summe der Bewertungen für die sechs Belastungskategorien berechnet, wobei eine Skala von 0 bis 10 verwendet wurde (siehe 7.2.3. „verwendete Fragebögen“). Dabei beschreibt 0 eine geringe und 10 eine hohe Belastung. Die Gesamtbeanspruchung für die Bearbeitung der Aufgabe 2 ergab folgende Werte:

Gesamtbeanspruchung Aufgabe 2			
	M	SD	N
XAI-System	15.9	9.1	10
Nicht-XAI-System	16.7	9.2	10

Tabelle 7: Vergleich der Gesamtbeanspruchung Aufgabe 2

Die folgende Abbildung 40 zeigt die einzelnen Belastungsarten im Vergleich:

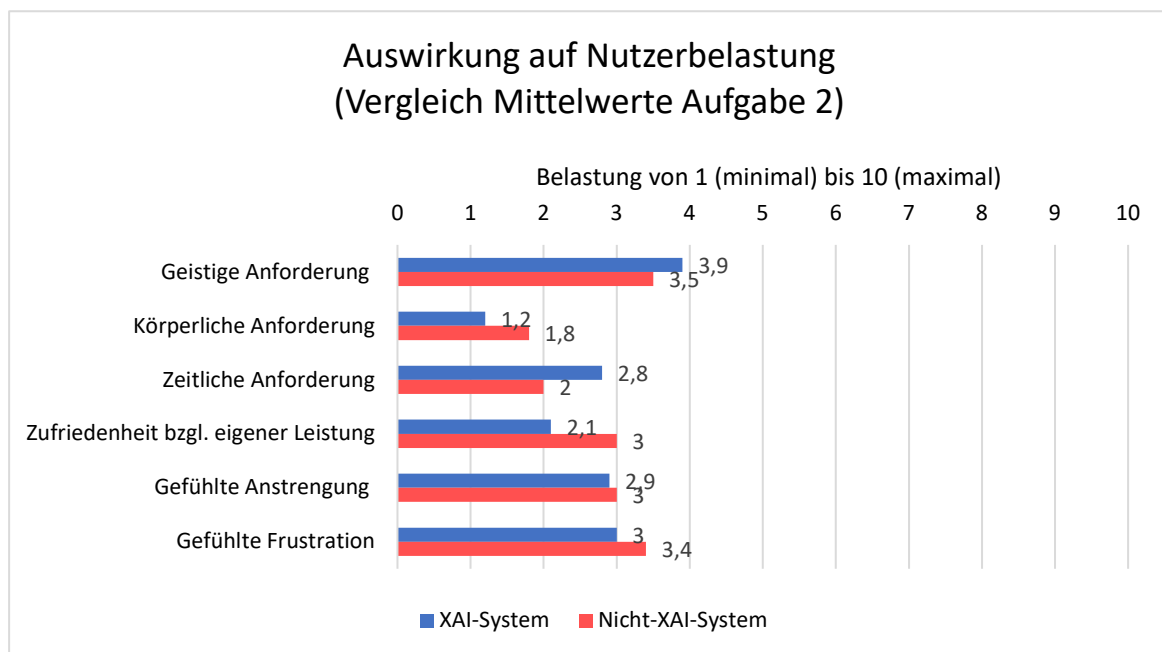


Abbildung 48: Vergleich der einzelnen Beanspruchungen bei Aufgabe 2, N=10

In Tabelle 8 werden die Mittelwerte und die Standardabweichungen für die einzelnen Belastungsarten bei Anwendung des XAI-Systems dargestellt:

<b>Ergebnisse zu den einzelnen Kategorien der Belastung für das XAI-System in Aufgabe 2</b>				
<b>Kategorie</b>	<b>M</b>	<b>SD</b>	<b>N</b>	
mentale Belastung	3.9	2.7	10	
körperliche Belastung	1.2	1.4	10	
zeitliche Belastung	2.8	2.7	10	
Zufriedenheit mit der eigenen Leistung	2.1	2.5	10	
Anstrengung	2.9	2.2	10	
Frustration	3.0	2.2	10	

*Tabelle 8: Einzelergebnisse für die Beanspruchung bei Anwendung des XAI-Systems*

### **Aufgabe 3: „Vorhersage der Risikobewertung“**

Die Gesamtbeanspruchung für die Bearbeitung der Aufgabe 3 ergab folgende Werte:

<b>Gesamtbeanspruchung Aufgabe 3</b>			
	<b>M</b>	<b>SD</b>	<b>N</b>
<b>XAI-System</b>	9.1	4.5	10
<b>Nicht-XAI-System</b>	21.6	11.5	10

*Tabelle 9: Vergleich der Gesamtbeanspruchung Aufgabe 3*



Die folgende Abbildung 41 zeigt die einzelnen Belastungsarten im Vergleich:

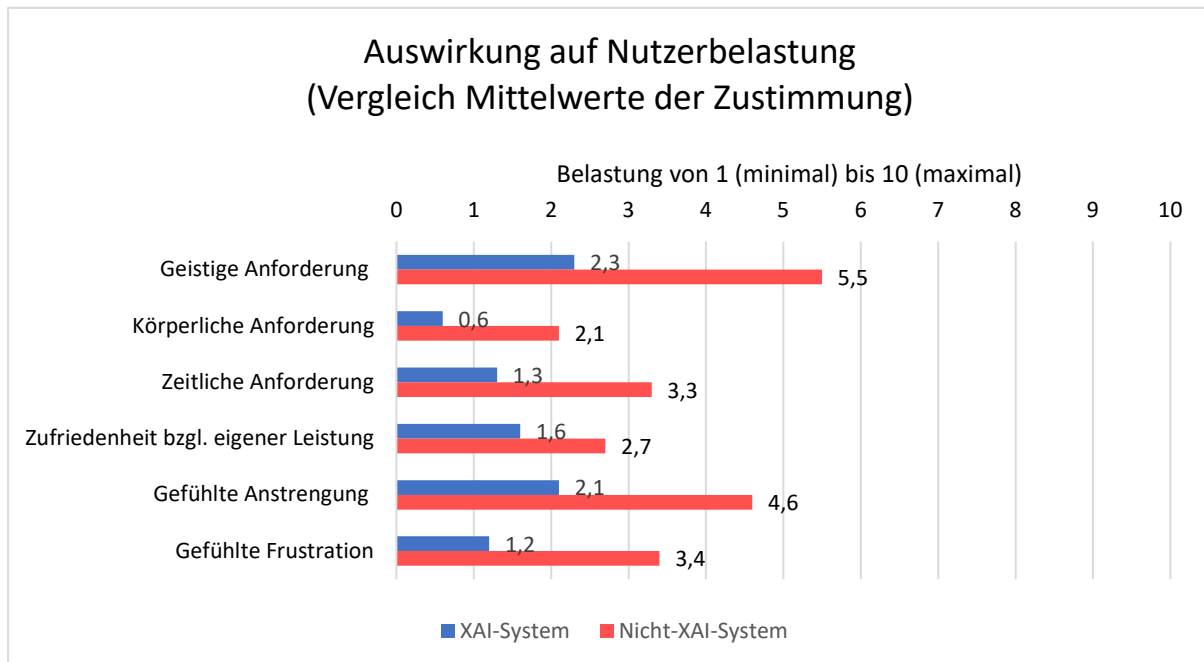


Abbildung 49: Vergleich der einzelnen Beanspruchungen bei Aufgabe 3, N=10

Die folgende Tabelle 10 stellt die Mittelwerte und Standardabweichungen der einzelnen Kategorien der Belastung bei Aufgabe 3 mit dem XAI-System dar:

Kategorie	M	SD	N
mentale Belastung	2.3	1.7	10
körperliche Belastung	0.6	0.7	10
zeitliche Belastung	1.3	0.9	10
Zufriedenheit mit der eigenen Leistung	1.6	1.2	10
Anstrengung	2.1	1.6	10
Frustration	1.2	0.7	10

Tabelle 10: Ergebnisse der einzelnen Beanspruchungen bei Aufgabe 3 mit XAI

### 7.3.3 Ergebnisse der Zeitmessung

Für den Vergleich der Zeit, die die Teilnehmer benötigten, um den Fehler zu finden, wurden ausschließlich die richtig beantworteten Ergebnisse miteinbezogen. Außerdem sind für Proband Nr. 2 die Ergebnisse für das Nicht-XAI-System verloren gegangen und werden mit NA angezeigt. Daher beziehen sich die Ergebnisse für die bewerteten Mittelwerte und Standardabweichungen in diesem Teil nicht auf N=10, sondern auf N=Teilnehmer, die richtig gelöst haben.

Außerdem werden folgende Abkürzungen verwendet, um die Ergebnisse der Aufgabenerfüllung zu beschreiben: Mit R ist die Anzahl der richtig und mit NR die Anzahl der nicht richtig abgegebenen Antworten gemeint. Das konnte ein falsch genannter Betrieb (Abgekürzt mit F), ein Abbruch der Aufgabe (Abgekürzt mit ABR) oder eine nicht verfügbare Antwort sein (Abgekürzt mit NA). N ist die Anzahl der Gesamtteilnehmer.

#### Ergebnisse der Zeitmessung für Aufgabe 2: "Modellfehler finden"

Der direkte Vergleich der Ergebnisse der durchschnittlichen Dauer aller richtigen Lösungen für beide Systeme zeigt, dass die benötigte Zeit nah beieinander liegt:

<i>Aufgabe 2 „Modellfehler finden“ - Vergleich der Mittelwerte der Dauer in Minuten</i>					
	<i>N</i>	<i>Anzahl TN, die richtig gelöst haben (R)</i>	<i>Anzahl TN, die nicht richtig gelöst haben (NR)</i>	<i>M</i>	<i>SD</i>
<b>XAI-System</b>	10	6	3 falsch, 1 Abbruch	8:46	3:01
<b>Nicht-XAI-System</b>	10	8	1 Abbruch, 1 NA	8:17	4:18

*Tabelle 11: Vergleich der Ergebnisse zur Bearbeitung von Aufgabe 2*

Beim Vergleich der Ergebnisse unter Hinzuziehung der Reihenfolge, welches System zuerst verwendet wurde, zeigen sich Unterschiede. Wird mit dem XAI-System begonnen, ist die durchschnittliche Dauer bei der Aufgabenerfüllung mit dem XAI-System größer als mit dem Nicht-XAI-System. Wird mit dem Nicht-XAI-System begonnen, ist die Dauer bei der Aufgabenerfüllung mit dem XAI-System kleiner als mit dem Nicht-XAI-System:

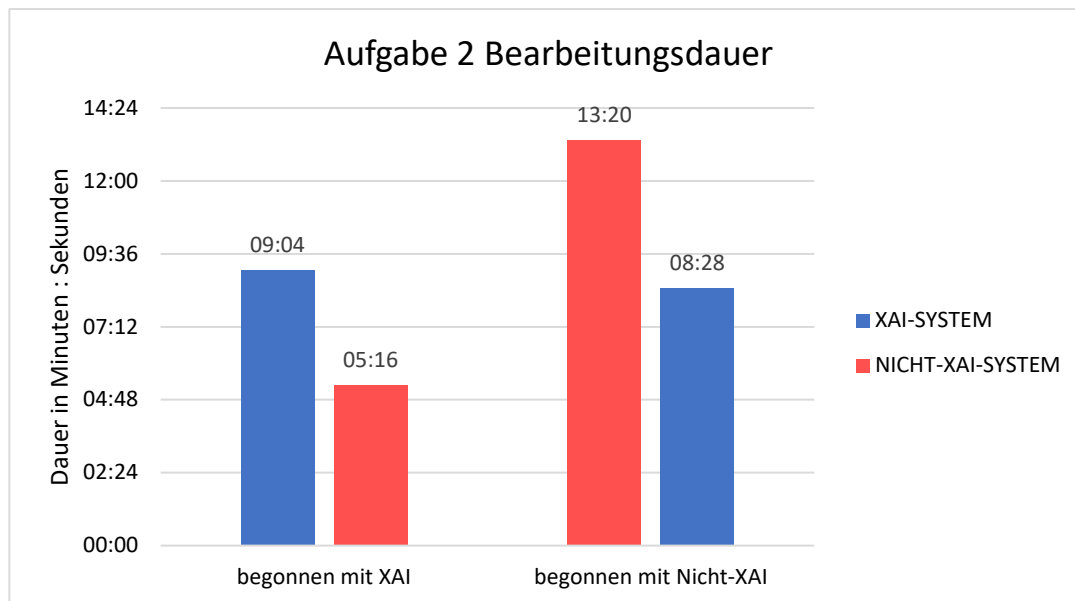


Abbildung 50: Vergleich der Bearbeitungsdauer Aufgabe 2

Tabelle 12 stellt die Ergebnisse zur Bearbeitung von Aufgabe 2 unter Berücksichtigung der Reihenfolge der Systeme dar.

bearbeitet mit / begonnen mit	XAI-System					Nicht-XAI-System				
	N	R	NR	M	SD	N	R	NR	M	SD
XAI-System	5	3	1 F 1 ABR	9:04	2:05	5	5	0	5:16	1:21
Nicht-XAI-System	5	3	2 F	8:28	2:32	5	3	1 ABR 1 NA	13:2 0	2:22

Tabelle 12: Vergleich Zeitmessung Aufgabe 2

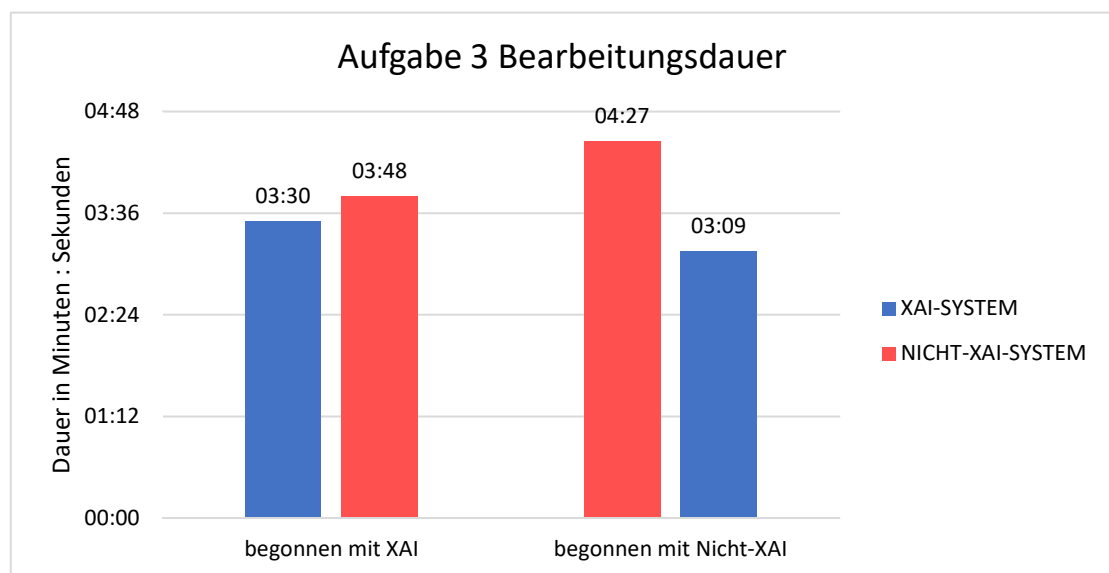
### Ergebnisse der Zeitmessung für Aufgabe 3: „Risikobewertung vorhersagen“

Beim Vergleich der durchschnittlichen Dauer für die Lösung der Aufgabe 3 (über alle richtigen Antworten hinweg, unabhängig von der Reihenfolge der Systeme) haben die Teilnehmer mit dem XAI-System weniger Zeit gebraucht.

<i>Aufgabe 3 - Risikowert schätzen - Vergleich der Mittelwerte der Dauer in Minuten</i>					
	<i>N</i>	<i>Anzahl TN, die richtig gelöst haben</i>	<i>Anzahl TN, die nicht richtig gelöst haben</i>	<i>M</i>	<i>SD</i>
<b>XAI-System</b>	10	8	2 falsch	3:22	4:18
<b>Nicht-XAI-System</b>	10	7	2 falsch, 1 NA	4:05	1:22

*Tabelle 13: Vergleich der Ergebnisse zur Bearbeitung von Aufgabe 3*

Wird die Reihenfolge zur Betrachtung hinzugezogen, zeigt sich, dass der Mittelwert bei der Aufgabenerfüllung mit dem XAI-System unter beiden Reihenfolgen kleiner ist als der Mittelwert bei der Aufgabenerfüllung mit dem Nicht-XAI-System:



*Abbildung 51: Vergleich der Bearbeitungsdauer Aufgabe 3*

Tabelle 14 stellt die Ergebnisse zur Bearbeitung von Aufgabe 3 unter Berücksichtigung der Reihenfolge der Systeme dar.

<i>bearbeitet mit / begonnen mit</i>	<i>XAI-System</i>					<i>Nicht-XAI-System</i>				
	<i>N</i>	<i>R</i>	<i>NR</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>R</i>	<i>NR</i>	<i>M</i>	<i>SD</i>
<b>XAI-System</b>	5	5	0	3:30	2:18	5	4	1F	3:48	1:37
<b>Nicht-XAI-System</b>	5	3	1F 1NA	2:26	2:26	5	3	1 F 1 NA	4:27	1:00

*Tabelle 14: Dauer in Minuten (Mittelwert) – Aufgabe 3 „Risikowert vorhersagen“*

### 7.3.4 Ergebnisse zur Gebrauchstauglichkeit (SUS-Werte)

Tabelle 12 stellt die Ergebnisse zur Gebrauchstauglichkeit beider Systemvarianten dar. Dabei bezeichnet Min den kleinsten gemessenen Wert und Max den größten in der Stichprobe:

	<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>	<b>N</b>
<b>XAI-System</b>	76.5	11.1	47.5	90	10
<b>Nicht-XAI-System</b>	59.3	12.3	45	82.5	10

*Tabelle 15: Vergleich der Ergebnisse der SUS-Abfrage*

Betrachtet man die Daten genauer, zeigt sich ein Ausreißer (A=47.5) im unteren Bereich bei Positionierung der unteren Ausreißerbegrenzung auf 2,5-mal Standardabweichung. Ohne Berücksichtigung des Ausreißers würde sich ein Mittelwert von 79,7 (SD = 6.1) bei N = 9 ergeben und die Werte zwischen 75 und 90 liegen.

### 7.3.5 Ergebnisse aus dem Quiz – Aufgabe 1 „Verständnis über das System“

Um nachvollziehen zu können, ob die Teilnehmer das System verstanden haben, wurde für Aufgabe 1 ein Quiz mit sechs Fragen zum System erstellt. Die folgende Tabelle stellt die Antworten dar:

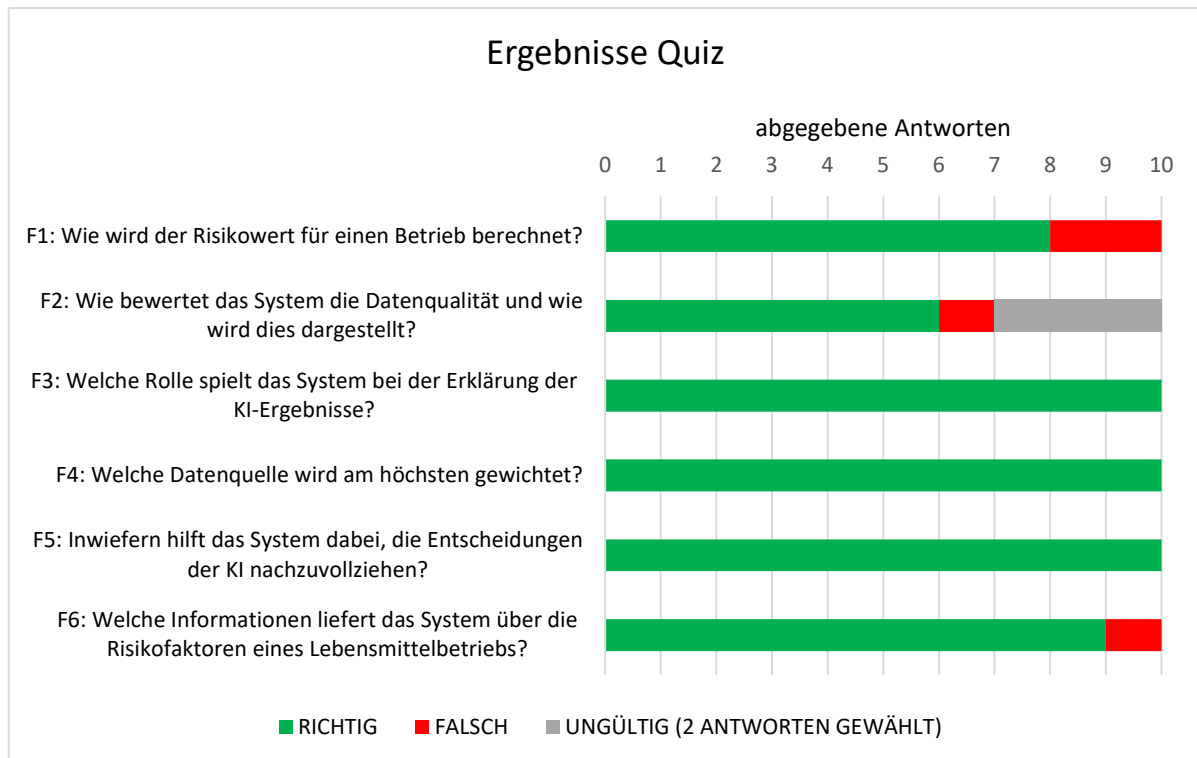


Abbildung 52: Quiz-Ergebnisse

- 50% der Verständnisfragen wurden von allen richtig beantwortet.
- 4 von 10 Teilnehmern haben alles richtig beantwortet
- 5 Teilnehmer haben eine Frage falsch beantwortet oder ungültig beantwortet
- 1 hat eine falsch und eine ungültig

Da jeder Teilnehmer einen Teil überwiegenden Teil der Fragen richtig beantwortet hat, musste niemand ausgeschlossen werden.

## 7.4 Diskussion

Im folgenden Abschnitt werden die im vorherigen Abschnitt präsentierten Ergebnisse diskutiert.

### 7.4.1 Vergleich der SIPA-Mittelwerte

Ein Vergleich der Gesamtmittelwerte über alle Aussagen zeigt, dass ein Angebot von Erklärungen ( $M=4.9$ ) einen deutlichen Einfluss auf die wahrgenommene Nachvollziehbarkeit eines Systems gegenüber einem System ohne Erklärungen ( $M=3.5$ ) hat. Auch bei der einzelnen Betrachtung der drei Schwerpunkte der SIPA-Skala findet das XAI-System mehr Zustimmung. Die Betrachtung der Item-Mittelwerte zeigt, dass das XAI-System durchschnittlich mit ca. einem oder mehr Skalenpunkten besser bewertet wurde als das Nicht-XAI-System. Besonders bei den Fragen zum „Verständnis des Ergebnisses des KI-Systems“ sind die Abstände mit 2 Skalenpunkten deutlich.

Interessant ist, dass bei dem Nicht-XAI-System die Probanden das Gefühl hatten, „die Informationen, die das System erfassen konnte“ seien für sie einsehbar gewesen ( $M=4.5$ ), obwohl ihnen bis auf die Rohdaten das System keine Informationen zur Informationsverarbeitung zur Verfügung gestellt hat. Für diese Auffälligkeit sind zwei Erklärungen denkbar. Einerseits ist möglich, dass die Nutzer in bestimmten Fällen Aspekte eines Systems als transparent wahrnehmen, selbst wenn keine expliziten Erklärungen gegeben werden. Den Nutzern kam nicht in den Sinn, dass ihnen zusätzliche Meta-Daten wie z.B. statistische Zusammenhänge, welche KI-Systeme für ihre Berechnungen verwenden könnten, verborgen geblieben sind. Andererseits könnte das Antwortverhalten der Probanden auch mit dem Versuchsdesign erklärt werden. Tatsächlich waren die Rohdaten im XAI-System und im Nicht-XAI-System quantitativ gleich. Ein Proband hatte also durchaus die Möglichkeit, alle eingeflossenen Rohdaten zu sichten. In einem echten KI-System wäre das nicht der Fall.

### 7.4.2 Vergleich der Nutzerbeanspruchung durch die Bearbeitung der Aufgaben

#### **Vergleich der Werte zur Beanspruchungshöhe bei Aufgabe 2**

Die Ergebnisse des Task Load Index zur Aufgabe „Modellfehler finden“ zeigen bei der Betrachtung der Gesamtbeanspruchung, dass die Aufgabenbewältigung mit dem XAI-System ( $M=15.9$ ) als ähnlich hohe Beanspruchung empfunden wurde wie die Arbeit mit dem Nicht-XAI-System ( $M=16.7$ ). Die Werte liegen damit im unteren Drittel des möglichen Wertebereichs. Die Belastung wurde also für beide Systemvarianten insgesamt als eher gering wahrgenommen. Der Mittelwert 15.9 für das XAI-System liegt knapp über dem untersten Quartil der mit dem TLX bewertbaren möglichen Belastungsbereich (max. 60).

Wie bereits im Kapitel 7.2.3 Abschnitt „verwendete Aufgabe“ vermutet wurde, hat sich die größere Menge an Interaktionsmöglichkeiten und Inhalten nicht positiv auf die empfundene

Gesamtbelastung ausgewirkt. Dies steht auch im Einklang mit Erkenntnissen aus der Literatur (siehe Kapitel 3.4).

Auch ein Vergleich der einzelnen Aussagen des Fragebogens zeigt, dass die spezifischen Anforderungen als ähnlich hoch empfunden wurden, da die Mittelwerte nur zwischen einem halben und einem ganzen Skalenpunkt auseinander liegen. Die mentale Beanspruchung war beim XAI-System leicht höher. Die Teilnehmer mussten deutlich mehr Inhalte verarbeiten und sahen das System zudem zum ersten Mal, ohne sich vorher einarbeiten zu können.

Auch die zeitliche Anforderung wurde beim XAI-System höher bewertet. Dies stimmt auch mit der Zeitmessung überein. Die Aufgabenbearbeitung dauerte durchschnittlich 30 Sekunden länger mit dem XAI-System.

Dagegen empfanden die Teilnehmer, dass sie mit dem XAI-System eine bessere Leistung erbracht haben, wie das Ergebnis bei dieser Frage mit etwa einen Skalenpunkt mehr im Vergleich zum Nicht-XAI-System zeigt.

Die empfundene mentale Belastung zeigte sich auch in den Aussagen der Teilnehmer in den Think-Aloud Daten. Viele Teilnehmer empfanden das viele Lesen der Datenquellen als unangenehm. Hier muss das System verbessert und das Betrachten der Daten so einfach wie möglich gestaltet werden. Statt den Zusammenfassungstext von oben nach unten lesen zu müssen, sollte ein Layout des Texts erstellt werden, der ein schnelles Überfliegen ermöglicht und wichtige Daten direkt sichtbar macht, sodass auch Nutzer mit wenig Lust zum Lesen nicht abgeschreckt sind.

Eine weitere Interpretationsmöglichkeit des Ergebnisses, ist die Aufgabengestaltung selbst. Bei dem Nicht-XAI-System konnten nur die Risikowerte verfälscht werden, da sonst nur noch die Rohdaten enthalten waren, von denen die Teilnehmer annehmen durften, dass diese der Wahrheit entsprechen. Im XAI-System wurde für diese Aufgabe zusätzlich zum Risikowert die „Zusammenfassung der Risikobewertung“ verfälscht, um es nicht zu einfach zu gestalten. Als ungeübte Nutzer fiel es den meisten Teilnehmern schwer, die relevanten Hinweise für den Fehler zu finden. Häufigere Aufgabendurchgänge und Zeit zum Kennenlernen der Anwendung würden das vermutlich verbessern. Dies gilt ebenso für das Nicht-XAI-System und die darin enthaltenen Rohdaten.

### **Vergleich der Werte zur Beanspruchungshöhe bei Aufgabe 3**

Ein anderes Bild ergibt sich beim Vergleich der Ergebnisse für den TLX-Fragebogen zur Aufgabe „Risikobewertung vorhersagen“. Die Gesamtbeanspruchung beim Nicht-XAI-System ( $M=21.6$ ) wird mehr als doppelt so groß empfunden im Vergleich zum XAI-System ( $M=9.1$ ). Auch die im Vergleich zur anderen Aufgabe und Systemvariante geringe Standardabweichung ( $SD=4,5$ ) deutet darauf hin, dass die Beanspruchung bei dieser Aufgabe über alle Teilnehmer hinweg konsistenter wahrgenommen wurde. Der deutliche Unterschied in der Gesamtbeanspruchung zwischen den Systemen deutet darauf hin, dass das XAI-System eine höhere Klarheit und Transparenz im



Entscheidungsprozess bietet und die XAI-Features wesentlich zur Reduzierung der Nutzerbeanspruchung beitragen konnten.

Bei Betrachtung der Mittelwerte der Einzelergebnisse fällt auf, dass die wahrgenommene mentale Anforderung beim XAI-System halb so viele Skalenpunkte groß ist, die Probanden eine deutlich geringere zeitliche Anforderung empfanden und zufriedener mit ihrer Leistung waren. Diese Ergebnisse unterstreichen die Wichtigkeit der Implementierung von XAI-Elementen in KI-Systemen, um die Nutzererfahrung zu verbessern.

### **Vergleich der Standardabweichungen bei den TLX-Fragebögen**

Der Vergleich der Standardabweichungen in den TLX-Ergebnissen bietet eine besondere Auffälligkeit. Die hohen Standardabweichungen in Ergebnissen der TLX-Abfragen für die Aufgabe 2 und der Abfrage für Aufgabe 3 mit dem Nicht-XAI-System ist bemerkenswert. Dies deutet darauf hin, dass die Wahrnehmung der Beanspruchung unter den Probanden stark variierte. Solch eine breite Streuung der Werte könnte auf unterschiedliche Faktoren wie individuelle Erfahrungen, Tagesform, Erwartungen oder persönliche Kompetenzen im Umgang mit den Systemen zurückzuführen sein.

### **7.4.3 Zusammenhang der Bearbeitungsdauer mit der Reihenfolge der Systemvariantennutzung**

Die durchschnittliche Dauer der Aufgabenbearbeitung, die für die Aufgabe „Modellfehler finden“ gemessen wurde, liegt für das XAI-System bei durchschnittlich 8 Minuten und 46 Sekunden ( $SD=3:01$ ) und mit dem Nicht-XAI-System bei 8 Minuten und 17 Sekunden ( $SD=4:18$ ) nah beieinander. Im direkten Vergleich ist die Bearbeitungszeit für das Nicht-XAI-System etwas kürzer als für das XAI-System, was darauf hindeuten könnte, dass das Nicht-XAI-System initial effizienter ist.

Wird die Reihenfolge mit welchem System begonnen wurde in Betracht gezogen, zeigt sich ein interessantes Muster: Die Bearbeitungsdauer mit dem zweiten/nachfolgenden System ist kürzer bei beiden Abfolgen. Dies deutet auf einen Lerneffekt hin, der bereits im Kapitel 7.2 „Design“ vermutet wurde, bei dem die Erfahrung mit einem System die Leistung im anderen System beeinflusst.

Wenn mit dem XAI-System begonnen wird, liegt die Dauer mit XAI bei 9:04, die Aufgabe wird anschließend mit Nicht-XAI (5:16) deutlich schneller bearbeitet. Es scheint, dass die Nutzung des XAI-Systems den Nutzern hilft, sich schneller in das Nicht-XAI-System einzuarbeiten. Umgekehrt, wenn mit Nicht-XAI (13:20) begonnen wird, dauert die Bearbeitung der Aufgabe in XAI im Vergleich länger (8:28). Dies könnte bedeuten, dass die Anfangserfahrung mit Nicht-Xai die Teilnehmer weniger effektiv auf die Bearbeitung der Aufgabe in System A vorbereitet. Zusammengefasst scheint es, dass die Erfahrung mit dem XAI-System den Nutzern einen Vorteil bietet, der sich positiv auf die Geschwindigkeit der Aufgabenbearbeitung im Nicht-XAI-System auswirkt.

Bei der Aufgabe „Risikowert vorhersagen“ brauchten die Teilnehmer mit dem XAI-System durchschnittlich weniger Zeit (3:22 gegenüber 4:05 im Nicht-XAI-System). Unabhängig von der Reihenfolge der Systemnutzung zeigte sich, dass die Leistung mit dem XAI-System besser war. Dies könnte darauf hindeuten, dass die Vorteile des XAI-Systems in dieser spezifischen Aufgabe stärker zum Tragen kommen.

Da allerdings mit der kleinen Stichprobengröße der einzelnen Gruppen (weiter vermindert durch die Herausnahme der nicht richtigen Antworten) die Stichprobe sehr klein ist, sollten diese Beobachtungen mit Vorsicht betrachtet werden.

#### 7.4.4 Vergleich der Ergebnisse zur Gebrauchstauglichkeit

Ein Vergleich der Mittelwerte aus den Ergebnissen des SUS-Fragebogens zeigt, dass die Gebrauchstauglichkeit des XAI-Systems ( $M=76.5$ ) deutlich höher bewertet wurde als die des Nicht-XAI-Systems ( $M=59.25$ ). Laut der Interpretationsskala nach Bangor (2009), erhält das XAI-System das Prädikat „Gut“, während das Nicht-XAI-System lediglich als „Ok“ eingestuft wird. Bei Entfernung eines unteren Ausreißers aus den Ergebnissen des XAI-Systems erreicht das System mit  $M=79.72$  fast die Schwelle zu „Exzellente“ ( $>80$ ).

Auch die Auswertung der Aussagen der Think-Aloud Methode zeigt keine großen Probleme in der Gebrauchstauglichkeit. Lediglich beim Schließen des Popup-Fensters zeigten sich teilweise unterschiedliche Vorstellungen vom Vorgehen: Ein Teilnehmer versuchte es mit der ESC-Taste zu schließen, während ein anderer außerhalb des Fensterbereichs klickte. Beide Teilnehmer haben den Close-Button aber direkt gefunden.

Der auffällige Ausreißer im XAI-System mit einem Wert von 47.5, weit entfernt vom Mittelwert, könnte verschiedene Gründe haben. Eine mögliche Erklärung könnte eine unzureichend präzise Präsentation des Systems und der Aufgaben sein. Dies wird durch die Beobachtung gestützt, dass die Teilnehmerin eine von zwei Teilnehmern war, die eine Aufgabe abgebrochen haben. Auch ihre SIPA-Bewertung für das XAI-System war mit 3.5 die niedrigste und gleich hoch bewertet wie das Nicht-XAI-System. Da sie in ihrem Beruf Richterin ist, könnte angenommen werden, dass das Lesen von Texten oder das Sichten von schriftlichen Daten nicht der Grund für eine eventuelle Überforderung gewesen ist. Für diese Annahme spricht auch, dass ihre Antwort für die gleiche Aufgabe mit dem Nicht-XAI-System richtig war und die kürzeste Bearbeitungsdauer hatte und das Nicht-XAI-System mit dem Satz kommentierte: „Hier muss man nur lesen“.

Abschließend lässt sich feststellen, dass der im Kapitel 7.2.3 (Abschnitt „verwendete Daten“) geäußerte Gedanke - mit der Abfrage des SUS-Werts, eine differenzierte Betrachtung zwischen Umsetzung der Anwendung als Interaktions- bzw. Navigationstool und der Anwendung als Informationsvermittler zu ermöglichen - wohl vermutlich mit diesem Untersuchungsdesign nicht umgesetzt werden konnte. Somit stellen die SUS-Ergebnisse vermutlich auch eine Bewertung der inhaltlichen Aspekte der Erklärungen bzw. der Gestaltung der Datensätze dar. Da das SUS-

Ergebnis für das XAI-System im guten Bereich liegt und auch die qualitativen Aussagen der Teilnehmer keine Probleme mit dem Interface zeigen, ist dies für die Interpretation des Ergebnisses weniger problematisch. Allerdings ist die Interpretation der SUS-Ergebnisse für das Nicht-XAI-System schwierig, da es nur als „Ok“ bewertet wurde, obwohl die Teilnehmer eine stark vereinfachte Version des XAI-Systems mit nur einem Button und einigen wenigen Dropdowns zu bedienen hatten. Eine zeitlich getrennte Untersuchung der Gebrauchstauglichkeit mit neutralen Daten könnte hilfreich sein, um die Ergebnisse von der Bewertung des Inhalts zu trennen.

#### 7.4.5 Limitierung

Die demografischen Daten der Studienteilnehmer zeigen, dass sie im Durchschnitt relativ jung und männlich sind ( $M = 30,4$  Jahre,  $SD = 5,7$  Jahre). In den Berufsfeldern dominieren akademische Berufe, von denen über der Hälfte im Bereich Informatik angesiedelt sind. Damit war die Teilnehmerzusammensetzung nicht ausreichend vielfältig, um eine Repräsentativität für die tatsächliche Zielgruppe des Systems darzustellen. Daher sind die Erkenntnisse aus dieser Evaluation nur eingeschränkt auf die breitere Nutzerbasis übertragbar.

Auch die begrenzte Teilnehmerzahl führt dazu, dass die Ergebnisse eher als vorläufige Hinweise zu betrachten sind, die durch umfangreichere und vielfältigere Studien validiert werden müssen, um verallgemeinerbare Erkenntnisse zu ermöglichen.

Da alle Teilnehmer aus dem persönlichen Umfeld des Erstellers kamen, kann sich in den Ergebnissen der Evaluation der Bias der Sozialen Erwünschtheit zeigen, was bedeutet, dass die Teilnehmer ihre Antworten möglicherweise so angepasst haben, um eine positive Wahrnehmung bei dem Ersteller der Studie zu erzeugen. Dies könnte zu einer Verzerrung der Ergebnisse führen, da die Antworten nicht unbedingt ihre wahren Meinungen widerspiegeln.

Außerdem ist eine Validierung, ob die Aufgaben das Intendierte messen, wie beispielsweise das Vorhersagen der Risikobewertung, erforderlich. Objektiv betrachtet haben die Teilnehmer in der Studie eingeschätzt, wie der Systementwickler den Betrieb bewertet. Diese Aufgabe muss mit echten Daten wiederholt werden, um ihre Validität zu bestätigen. Ebenso ist die Aufgabe des Modellfehlerfindens nicht klar genug spezifiziert. Es stellt sich die Frage, auf welcher Ebene der Systemarchitektur (siehe Kapitel 4.6) dieser Fehler eigentlich stattgefunden hat. Dies hat sich auch in den unterschiedlichen Erwartungen der Teilnehmer an den Fehler widerspiegelt.

#### 7.5 Fazit Evaluation

Die Evaluation dieser Arbeit konzentriert sich auf die Überprüfung eines User-Interfaces, das nach den Prinzipien des Explainable Artificial Intelligence (XAI) gestaltet wurde. Das Hauptziel ist es, festzustellen, ob dieses Interface zu einer höheren Nachvollziehbarkeit der KI-Ergebnisse bei den Nutzern führt. Drei Kriterien sind dafür maßgebend: Transparenz, Verständnis der KI-

Ergebnisse und Vorhersagbarkeit dieser Ergebnisse. Ein weiteres Ziel ist es, die Gebrauchstauglichkeit zu messen.

Die methodische Herangehensweise der Evaluation umfasst ein multi-methodisches Design. Sie nutzt einen A/B-Test, bei dem zwei Systemvarianten (mit und ohne XAI) gegeneinander getestet werden, um die Effekte von XAI auf die Nachvollziehbarkeit zu messen. Die Teilnehmer durchlaufen beide Bedingungen der Studie in einem Within-Subjects-Design, wobei die Reihenfolge der Bedingungen randomisiert wurde. Zusätzlich wird die Think-Aloud-Methode angewandt, um qualitative Daten zu generieren und ein tieferes Verständnis für die Interaktionen und Bewertungen der Teilnehmer zu erhalten. Eine Mindestanzahl von zehn Teilnehmern konnte erreicht werden.

Verschiedene Aufgaben wurden konzipiert, um die Metriken zu erfassen und sowohl das System objektiv zu bewerten als auch eine subjektive Selbstbewertung der Erfahrungen und Wahrnehmung der Teilnehmer zu ermöglichen. Diese Aufgaben basieren auf den operationalisierten Kriterien von Transparenz, Verständnis und Vorhersagbarkeit. Zusätzlich werden verschiedene Fragebögen eingesetzt, darunter der SUS-Fragebogen für die Bewertung der Usability und der SIPA-Fragebogen für die Bewertung der Nachvollziehbarkeit.

Die Ergebnisse der Studie zeigen, dass das XAI-System (System A) insgesamt eine höhere Nachvollziehbarkeit und Gebrauchstauglichkeit aufweist als das Nicht-XAI-System (System B). Dies deutet auf eine allgemein bessere Nutzererfahrung beim XAI-System hin.

Im Bereich der Aufgabenbearbeitung (TLX-Ergebnisse) zeigt sich, dass für die Aufgabe „Modellfehler finden“ das XAI-System eine ähnliche Gesamtbeanspruchung wie das Nicht-XAI-System aufweist, jedoch das Nicht-XAI-System eine geringere mentale Belastung verursacht. Die Gesamtbeanspruchung war bei der ersten Aufgabe (Modellfehler finden) für das XAI-System etwas niedriger als für das Nicht-XAI-System, und bei der zweiten Aufgabe (Risikobewertung schätzen) deutlich niedriger.

Hinsichtlich der Zeitmessung waren die Ergebnisse für die Aufgabe 2 zwischen den beiden Systemen ähnlich, mit leichten Vorteilen für das XAI-System. Es zeigt sich aber ein Lerneffekt, da die Dauer der Aufgabenbearbeitung mit dem zweiten System kürzer war, unabhängig von der Reihenfolge der Systemnutzung. Bei der Aufgabe 2 waren die Teilnehmer stets mit dem XAI-System schneller.

Zusammenfassend lässt sich sagen, dass das XAI-System im Vergleich zum Nicht-XAI-System in Bezug auf die Nachvollziehbarkeit, die Beanspruchung bei der Aufgabenbearbeitung und die Gebrauchstauglichkeit besser abschneidet. Diese Ergebnisse unterstreichen die Bedeutung von Erklärbarkeit und Transparenz in KI-Systemen für die Nutzererfahrung.

## 8 Fazit

Im abschließenden Kapitel dieser Arbeit wird zunächst ein Fazit (8.1) gezogen sowie offene Punkte angesprochen (8.2). Zu allerletzt wird ein Ausblick gegeben, inwiefern das Thema der Arbeit weiter erforscht werden könnte.

### 8.1 Zusammenfassung

In dieser Arbeit wurde ein Interface für die Planung von Lebensmittelkontrollen mit Hilfe eines Entscheidungsunterstützungssystem entwickelt. Das System macht Lebensmittelkontrolleuren Vorschläge, welche Betriebe sie als nächstes kontrollieren sollten. Diese Vorschläge basieren auf KI-generierten Risikobewertungen der zu kontrollierenden Betriebe.

Das entwickelte System soll die Nutzer dabei unterstützen, die Berechnung der Risikobewertung und dahinterliegenden Prozesse nachvollziehen zu können. Zu diesem Zweck stellt das System Erklärungen und Begründungen in verschiedenen Formaten bereit, die die Entscheidungsfindung des KI-Systems und die beteiligten Prozesse transparent machen sollen.

Da die Entscheidung, welche Betriebe kontrolliert werden, einen Entscheidungsspielraum umfasst, darf sie nicht allein von einem KI-System getroffen werden. Die endgültige Entscheidung liegt daher beim Kontrolleur für die er sogar u.U. haftbar ist. Darum ist es zentral, dass den Kontrolleuren die Gründe, warum ein Betrieb kontrolliert werden sollte, in einer Form präsentiert werden, die für sie nachvollziehbar ist. Daher ist eine nutzerzentrierte Gestaltung des Systems wichtig.

Für die Umsetzung dieses Ziels wurden zwei Bereiche betrachtet. Zum einen wurde der Forschungsbereich Explainable AI betrachtet, der sich damit beschäftigt, wie Nutzern die Ergebnisse und Prozesse eines KI-Systems verständlich erklärt werden können. Zum anderen wurde der Nutzungskontext der Lebensmittelüberwachung analysiert, in dem die Aufgaben, die Nutzer und die verwendete Software betrachtet wurden. Dafür wurde u.a. die Behörde der Lebensmittelüberwachung Kiel besucht und ein Nutzer- bzw. Experteninterview geführt. Aus den Ergebnissen dieser Recherche wurden Anforderungen an das System abgeleitet. Diese Anforderungen wurden dann in Funktionalitäten für das System übersetzt. Auf dieser Grundlage wurde in zwei formativ evaluierten Iterationen ein Gestaltungskonzept erstellt, das schließlich als High-Fidelity-Prototyp umgesetzt wurde.

Um zu überprüfen, ob die Ziele der Arbeit erfüllt wurden, wurde in einer abschließenden Evaluation folgende Frage beantwortet: „Ist ein System, das auf Gestaltungsprinzipien für Explainable AI entwickelt wurde, nachvollziehbarer als ein System, das nicht auf solchen Prinzipien basiert?“

Um diese Frage zu beantworten, wurde eine Versuchsstudie mit einem A/B-Test entwickelt. Dafür wurden aus dem entwickelten Prototyp zwei Varianten erstellt: Ein XAI-System und ein System ohne XAI-Feature.

Durch die Operationalisierung von Nachvollziehbarkeit nach Schrills et al. und ihrer entwickelten SIPA-Skala, wurde die wahrgenommene Nachvollziehbarkeit erhoben. Darüber hinaus wurden Aufgaben konzipiert, die es ermöglichten, verschiedene Metriken zu messen, sodass deren Bearbeitung mit dem XAI-System zusätzlich objektive Daten lieferten und weitere Daten ermöglichten. So konnte mittels des NASA-TLX-Fragebogens auch die Beanspruchung der Nutzer durch das System gemessen werden und die Bearbeitungsdauern verglichen werden. Außerdem wurde die Gebrauchstauglichkeit mittels des SUS-Fragebogen abgefragt.

Die Auswertung der Daten lässt darauf schließen, dass die für die Evaluation gestellte Frage, ob ein auf XAI-basiertes System die Ergebnisse eines KI-Systems nachvollziehbarer macht als ein System ohne XAI-Feature, positiv beantwortet werden kann. Sowohl in Bezug auf die subjektive Wahrnehmung als auch objektive Maße (Bearbeitungsdauer) kann XAI eine Unterstützung sein. Auch die Auswertung der Daten zur Gebrauchstauglichkeit zeigen, dass der Prototyp diesbezüglich positiv abschneidet. Damit konnte das Ziel der Evaluation und das Primärziel der Arbeit, eine nachvollziehbare Darstellung einer KI-basierten Bewertung von Lebensmittelbetrieben, erfüllt werden.

Das Angebot von Erklärungen bietet für Erstnutzer jedoch nicht unbedingt eine Entlastung bezüglich der mentalen Anforderung und der Bearbeitungsdauer einer Aufgabe. Es kommt sowohl auf die Aufgabe als auch auf die Wahrnehmung des individuellen Nutzers an. Entscheidend ist außerdem die Form, in der die Erklärungen präsentiert werden. Diese sollten in unterschiedlichen Modalitäten präsentiert werden, um die verschiedenen Bedürfnisse der verschiedenen Nutzer zu befriedigen.

## 8.2 Offene Punkte

Aus der Liste der Anforderungen an das System wurden die wichtigsten Punkte für diese Arbeit umgesetzt. Die noch offenen Punkte sollen im Folgenden bezüglich nicht umgesetzter Funktionslogik und inhaltlichen Aspekten unterschieden werden.

### **Funktionen**

Zwei Anforderungen mit der Priorisierung „Muss“, die die Funktionalität der Planung der Kontrollen betreffen, wurden nur statisch umgesetzt. Dies sind die Datumseingabe und die Ausgabe der Liste. Es wurde keine Logik für diese Funktionen implementiert. Dies liegt vor allem daran, dass im Laufe der Arbeit klar wurde, dass keine große Anzahl Datensätze produziert werden können, sodass eine dynamische Veränderung, beispielsweise durch Eingabe unterschiedlicher Zeiträume eine Ausgabe unterschiedlicher Betriebe, nicht zielführend ist. Diese Funktionalität wurde letztendlich für die Evaluation nicht benötigt.

Zwei Anforderungen mit der Priorisierung „Kann“, die nicht umgesetzt wurden, sind die Kartenansicht und der Profilbereich. Auch bei diesen beiden Anforderungen spielen die fehlenden Daten eine Rolle. Eine Kartenansicht kann hilfreich sein, wenn damit Informationen dargestellt werden können. Diese Informationen haben für diese Arbeit gefehlt bzw. konnten nicht produziert werden. Der Profilbereich hätte für die personalisierte Anpassbarkeit der Erklärungen nützlich sein können, für die es aber ebenso keine Datensätze gab.

### **Inhalt**

Eine Anforderung, die die Inhalte betrifft und mit „kann“ priorisiert wurde, ist die angesprochene Anpassbarkeit der Sprache der Erklärttexte durch die Nutzer. Auch hier hat es an Inhalten bzw. Zeit für die Umsetzung gefehlt.

Eine Anforderung bei der nicht eindeutig ist, ob diese erfüllt wurde, ist, dass die Zusammenfassungen der Bewertung kurzgehalten sein sollte. Die Länge der Texte lag bei 250-280 Zeichen. Dies war für einige Evaluationsteilnehmer zu lang. Es könnte von der individuellen Leseaffinität der Nutzer abhängen. Eine Verbesserung des Layouts bzw. der Übersichtlichkeit der Zusammenfassung sollte jedoch umgesetzt werden.

### **Evaluation der Nachvollziehbarkeit**

Ein weiterer offener Punkt, der sich zwar nicht unbedingt aus den Anforderungen ergibt, sondern aus dem grundsätzlichen Thema der Arbeit, ist die Einordnung der wahrgenommenen Nachvollziehbarkeit des Systems. In der Evaluation wurde diese Variabel zwar erhoben, aber nur vergleichend zu einem anderen System. Das Ergebnis aus der Evaluation ließ keine Aussage darüber zu, wie hoch die Nachvollziehbarkeit ist. Folglich konnte in der Arbeit nicht festgestellt werden, ob das entwickelte System tatsächlich in besonderem Maße als nachvollziehbar wahrgenommen wurde.

## **8.3 Ausblick**

Über die in diesem Rahmen geleistete Arbeit und die noch offenen Punkte hinaus, sind noch weitere mögliche Untersuchungen interessant. Der Ausblick soll als Ausweitung eines gedanklichen Rahmens ausgehend von der entwickelten Arbeit betrachtet werden.

### **Innerhalb des bereits Umgesetzten**

Die in der Evaluation umgesetzten statistischen Methoden gingen nur wenig über den Vergleich von Mittelwerten hinaus. Es ist denkbar, dass mit den bereits erhobenen Daten und der richtigen Wahl statistischer Tests weitere Ergebnisse generiert werden können. Beispielsweise konnte in dieser Arbeit keine Korrelationsanalyse umgesetzt werden, um die Beziehungen zwischen verschiedenen Variablen (z.B. zwischen SIPA-, SUS- und NASA-TLX-Scores und den Zeitwerten) zu untersuchen. Auch ein gepaarter T-Test für den Vergleich der Mittelwerte ist denkbar, um zu bestimmen, ob die beobachteten Unterschiede zwischen den Gruppen statistisch signifikant sind.

### **Nach Umsetzung der offenen Punkte und Erkenntnissen der Evaluation**

Nach Umsetzung der Erkenntnisse aus der Evaluation zur Verbesserung des Systems in einer weiteren Iteration, wäre es interessant zu wissen, wie die Ergebnisse ausfallen, wenn die Evaluationssteilnehmer das System und den Kontext besser kennen würden. Ein Untersuchungsdesign mit mehr Zeit für das Kennenlernen des Systems und der Daten und die Durchführung mehrerer Durchgänge könnte, bereits mit diesem Prototyp weitere Einsichten in die Gestaltung dieses Systems liefern. In diesem Szenario wäre das Layout der fiktiven Datensätze verbessert und könnte damit eventuell deutlicher die Unterschiede zwischen einem XAI- und einem Nicht-XAI-System darstellen.

### **Weitere Analyse des Nutzungskontextes und Präzisierung der Anforderungen**

Noch spannender wäre die Umsetzung des Prototyps mit echten Daten und mit zusätzlicher Expertise von Lebensmittelkontrolleuren sowohl während der Gestaltung als auch in der abschließenden Evaluation. Bezüglich der Daten wäre z.B. die Frage nach individuell anpassbaren Parametern für Berechnung der KI-Ergebnisse, beispielsweise für die Gewichtung der Daten, interessant. Weitere Experteninterviews mit Lebensmittelkontrolleuren und XAI-Experten könnten helfen, die Fragen zu beantworten, welche Informationen die Nutzer wünschen oder welche sie aus Expertensicht brauchen. Das Bestimmen der konkreten Erklärungsbedürfnisse der Lebensmittelkontrolleure sollte ein wichtiger Aspekt der weiteren Gestaltung sein. Mit echten Daten ließe sich außerdem herausfinden, welche non-verbalen Visualisierungen der Daten im Kontext der Lebensmittelüberwachung möglich sind. Die in der ersten Gestaltungsiteration dieser Arbeit (Kapitel 4.5.1) noch verfolgte Idee der Zusammenfassungsansicht des KI-Ergebnisses als multi-modales Dashboard mit zusätzlichen Visualisierungen könnte möglicherweise verwirklicht werden.

Der Ausblick zeigt, dass es lohnenswert wäre, das Thema Nachvollziehbarkeit und XAI weiter zu beforschen. Besonders in sensiblen Kontexten, wie Behörden, ist Nachvollziehbarkeit von KI-Systemen unverzichtbar, um sie in diesen Bereichen verstärkt nutzen zu können. KI-Systeme bieten große Potenziale für eine effektive und effiziente öffentliche Verwaltung, aber nur wenn sie von ihren Nutzern bewusst verwendet werden.



# Abbildungsverzeichnis

Abbildung 1: Menschenzentrierter Gestaltungsprozess (DIN EN ISO 9241-210).....	3
Abbildung 2: Ziele für Erklärbarkeit, die von versch. Zielgruppen angestrebt werden.....	12
Abbildung 3: Design Framework für die Kommunikation von KI-Systemen.....	16
Abbildung 4: Ergebnis des Ideensammelns beim Workshop.....	29
Abbildung 5: Datenverarbeitung des fiktiven KI-Systems.....	30
Abbildung 6: Listenansicht.....	36
Abbildung 7: Kartenansicht.....	37
Abbildung 8: Ansicht der Zusammenfassung der Bewertung.....	38
Abbildung 9: Gewichtungsansicht als Balken- & Tortendiagramm.....	38
Abbildung 10: Rohdaten.....	39
Abbildung 11: Datenqualität.....	40
Abbildung 12: Von Atoms zu Molecules.....	42
Abbildung 13: Farben für die Interfacegestaltung.....	43
Abbildung 14: Farbspektrum für die Risikobewertung.....	43
Abbildung 15: Mockup der Listenansicht.....	44
Abbildung 16: Mockup Kartenansicht.....	45
Abbildung 17: Mockup zur Ansicht der Zusammenfassung der Erklärung.....	46
Abbildung 18: Mockup Gewichtungsansicht.....	46
Abbildung 19: Mockup der Rohdatenansicht.....	47
Abbildung 20: Schritte vom Dateninput zum User.....	49
Abbildung 21: Code-Struktur.....	54
Abbildung 22: Import der JSON-Datei.....	55
Abbildung 23: entries Variable.....	55
Abbildung 24: Methode zum Filtern des JSON Inhalts.....	56
Abbildung 25: Weitergabe von Daten an Unterkomponenten.....	56
Abbildung 26: content Props.....	56
Abbildung 27: Ausschnitt des Datensatzes in der JSON-Datei.....	57
Abbildung 28: Risikoindikator.....	58
Abbildung 29: Methode zur Bestimmung des Farbwerts.....	58
Abbildung 30: Split-Button.....	59
Abbildung 31: Dropdown des Split-Buttons.....	59
Abbildung 32: Implementierung des Split-Buttons.....	60
Abbildung 33: Klick-Events für die Wahl der Erkläransichten.....	60
Abbildung 34: Übergabe Props an Chart.js.....	60
Abbildung 35: Konfiguration des summierten Balkendiagramms.....	61
Abbildung 36: die zwei Typen Balkendiagramme der Gewichtungsansicht.....	61
Abbildung 37: dataset Konfiguration für stacked bar charts.....	62
Abbildung 38: Konfiguration der speziellen Labelanzeige in Chart.js.....	62
Abbildung 39: Implementierte Listenansicht.....	64
Abbildung 40: Implementiertes Dropdown-Menü.....	64
Abbildung 41: Implementierte Gewichtungsansicht.....	65
Abb. 42: Implementierte Einzelbewertung Dropdown.....	65
Abbildung 43: Implementierte Miniaturansicht.....	66
Abbildung 44: Implementierte Rohdatenansicht.....	66
Abbildung 45: Schließen der vergrößerten Miniaturansicht.....	67
Abbildung 46: Schließen des Popups.....	67

Abbildung 47: SIPA Ergebnisse beider Systeme, N=10 .....	78
Abbildung 48: Vergleich der einzelnen Beanspruchungen bei Aufgabe 2, N=10 .....	80
Abbildung 49: Vergleich der einzelnen Beanspruchungen bei Aufgabe 3, N=10 .....	82
Abbildung 50: Vergleich der Bearbeitungsdauer Aufgabe 2 .....	84
Abbildung 51: Vergleich der Bearbeitungsdauer Aufgabe 3 .....	85
Abbildung 52: Quiz-Ergebnisse .....	87

## Tabellenverzeichnis

Tabelle 1: Ermittelte Faktoren für die Nachvollziehbarkeit eines Systems .....	20
Tabelle 2: Anforderungen und Funktionalitäten .....	33
Tabelle 3: Liste der Evaluationsteilnehmer .....	71
Tabelle 4: Vergleich Gesamtmittelwerte für die Erhebung der Nachvollziehbarkeit .....	77
Tabelle 5: Ergebnisse Einzelaussagen zur Nachvollziehbarkeit (XAI-System) .....	79
Tabelle 6: Ergebnisse Einzelaussagen zur Nachvollziehbarkeit (Nicht-XAI-System) .....	79
Tabelle 7: Vergleich der Gesamtbeanspruchung Aufgabe 2 .....	80
Tabelle 8: Einzelergebnisse für die Beanspruchung bei Anwendung des XAI-Systems .....	81
Tabelle 9: Vergleich der Gesamtbeanspruchung Aufgabe 3 .....	81
Tabelle 10: Ergebnisse der einzelnen Beanspruchungen bei Aufgabe 3 mit XAI .....	82
Tabelle 11: Vergleich der Ergebnisse zur Bearbeitung von Aufgabe 2 .....	83
Tabelle 12: Vergleich Zeitmessung Aufgabe 2 .....	84
Tabelle 13: Vergleich der Ergebnisse zur Bearbeitung von Aufgabe 3 .....	85
Tabelle 14: Dauer in Minuten (Mittelwert) – Aufgabe 3 „Risikowert vorhersagen“ .....	86
Tabelle 15: Vergleich der Ergebnisse der SUS-Abfrage .....	86

## Literaturverzeichnis

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>

*Automatisierte Entscheidungssysteme im öffentlichen Sektor.* (2022, Februar 25). AlgorithmWatch. <https://algorithmwatch.org/de/adm-offentlichersektor-empfehlungen/>

*AVV Rahmen-Überwachung – AVV RÜb.* (2021, Januar 20). [https://www.verwaltungsvorschriften-im-internet.de/bsvwvbund\\_20012021\\_3158100140002.htm](https://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_20012021_3158100140002.htm)

Bangor, A. (2009). *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale.* 4(3).

Bernnat, R., Halsch, V., Mette, P., Linder, A., Stindt, C., Daub, L., & Schmidtke, H.-C. (2022, Juni). *Fachkräftemangel im öffentlichen Sektor | Zehn Handlungsempfehlungen für Entscheider:innen.* PwC. <https://www.pwc.de/de/branchen-und-markte/oeffentlicher-sektor/fachkraeftemangel-im-oeffentlichen-sektor.html>

Born, C. (2020, Mai 13). *Personalmangel und Coronakrise—"Lebensmittelkontrollen bedeuten aktiven Gesundheitsschutz".* Deutschlandfunk. <https://www.deutschlandfunk.de/personalmangel-und-coronakrise-lebensmittelkontrollen-100.html>

Brooke, J. (1995). *SUS - A quick and dirty usability scale.*

BVLK. (2022). *Was machen Lebensmittelkontrolleure?*

Chromik, M., & Butz, A. (2021). Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Hrsg.), *Human-Computer Interaction – INTERACT 2021* (S. 619–640). Springer International Publishing. [https://doi.org/10.1007/978-3-030-85616-8\\_36](https://doi.org/10.1007/978-3-030-85616-8_36)

Ehsan, U., & Riedl, M. O. (2022). *Social Construction of XAI: Do We Need One Definition to Rule Them All?* (arXiv:2211.06499). arXiv. <https://doi.org/10.48550/arXiv.2211.06499>

Engelmann, J., & Puntschuh, M. (2020). *KI im Behördeneinsatz: Erfahrungen und Empfehlungen.*

*Ethics guidelines for trustworthy AI.* (2019, April 8). <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Etscheid, J., Lucke, J., & Stroh, F. (2020). *KÜNSTLICHE INTELLIGENZ IN DER ÖFFENTLICHEN VERWALTUNG | Anwendungsfelder und Szenarien.* <http://publica.fraunhofer.de/dokumente/N-577708.html>

Fischer, S., & Petersen, T. (2018). Was Deutschland über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage. *Bertelsmann Stiftung - Impuls Algorithmenethik.* <https://doi.org/10.11586/2018022>

- Foodwatch. (2019). *Jede dritte vorgeschriebene Lebensmittelkontrolle in Deutschland fällt aus – Datenrecherche zu 400 Behörden in ganz Deutschland offenbart politisches Versagen beim Verbraucherschutz*. <https://www.foodwatch.org/de/pressemitteilungen/2019/jede-dritte-vorgeschriebene-lebensmittelkontrolle-in-deutschland-faellt-aus/?cookieLevel=not-set>
- Haunhorst, E. (2019, November 5). *Verbraucherschutz zu Lebensmittelskandal— "Wir müssen intelligenter kontrollieren"*. Deutschlandfunk. <https://www.deutschlandfunk.de/verbraucherschutz-zu-lebensmittelskandal-wir-muessen-100.html>
- Herczeg, M. (2009). Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme, 3. Auflage. In *Software-Ergonomie*. <https://doi.org/10.1515/9783110446869>
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, 5. <https://www.frontiersin.org/articles/10.3389/fcomp.2023.1096257>
- Holzinger, A. (2021). The Next Frontier: AI We Can Really Trust. In M. Kamp, I. Koprinska, A. Bibal, T. Bouadi, B. Frénay, L. Galárraga, J. Oramas, L. Adilova, Y. Krishnamurthy, B. Kang, C. Largeron, J. Lijffijt, T. Viard, P. Welke, M. Ruocco, E. Aune, C. Gallicchio, G. Schiele, F. Pernkopf, ... G. Graça (Hrsg.), *Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (S. 427–440). Springer International Publishing. [https://doi.org/10.1007/978-3-030-93736-2\\_33](https://doi.org/10.1007/978-3-030-93736-2_33)
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4), e1312. <https://doi.org/10.1002/widm.1312>
- Kamath, U., & Liu, J. (2021). *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-83356-5>
- Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2022). How to explain AI systems to end users: A systematic literature review and research agenda. *Internet Research*, 32(7), 1–31. <https://doi.org/10.1108/INTR-08-2021-0600>
- Laubheimer, P. (2019, Mai 12). *Split Buttons: Definition*. Nielsen Norman Group. <https://www.nngroup.com/articles/split-buttons/>
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Chapter 1—Introduction to HCI research. In J. Lazar, J. H. Feng, & H. Hochheiser (Hrsg.), *Research Methods in Human Computer Interaction (Second Edition)* (S. 1–24). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-805390-4.00001-7>
- Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3313831.3376590>
- Liao, Q. V., & Varshney, K. R. (2022). *Human-Centered Explainable AI (XAI): From Algorithms to User Experiences* (arXiv:2110.10790). arXiv. <https://doi.org/10.48550/arXiv.2110.10790>
- Lopes, P., Silva, E., Braga, C., Oliveira, T., & Rosado, L. (2022). XAI Systems Evaluation: A Review of Human and Computer-Centred Methods. *Applied Sciences*, 12(19), 9423. <https://doi.org/10.3390/app12199423>

- Martini, M., & Nink, D. (2018). Subsumtionsautomaten ante portas? – Zu den Grenzen der Automatisierung in verwaltungsrechtlichen (Rechtsbehelfs-) Verfahren. *Deutsches Verwaltungsblatt*, 133(17), 1128–1137. <https://doi.org/10.1515/dvbl-2018-1331707>
- Miller, T. (2018). *Explanation in Artificial Intelligence: Insights from the Social Sciences* (arXiv:1706.07269). arXiv. <https://doi.org/10.48550/arXiv.1706.07269>
- NASA TLX. (2022, Juni 15). <https://humansystems.arc.nasa.gov/groups/TLX/tlxpaperpencil.php>
- Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann Publishers Inc.
- Nielsen, J. (2000, März 18). *Why You Only Need to Test with 5 Users*. Nielsen Norman Group. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Overdiek, M., & Petersen, T. (2022, April). *Was Deutschland über Algorithmen weiß und denkt*. Bertelsmann Stiftung. <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/was-deutschland-ueber-algorithmen-weiss-und-denkt>
- Rummel, B. (2013). *System Usability Scale (Translated into German)*.
- Schrills, T., & Franke, T. (2022). *How do Users Experience Traceability of AI Systems? Examining Subjective Information Processing Awareness in Automated Insulin Delivery (AID) Systems* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/3v9b8>
- Schrills, T., Zoubir, M., Bickel, M., Kargl, S., & Franke, T. (2021, Mai 8). *Are Users in the Loop? Development of the Subjective Information Processing Awareness Scale to Assess XAI*.
- Suresh, H., Gomez, S. R., Nam, K. K., & Satyanarayan, A. (2021). Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3411764.3445088>
- Verbraucherbeschwerde*. (o. J.). BVLK. Abgerufen 1. Dezember 2023, von <https://bvlk.de/verbraucherbeschwerde.html>
- Vollständig automatisierter Erlass eines Verwaltungsaktes, § 35a Verwaltungsverfahrensgesetz (Bund) § Teil III - Verwaltungsakt (§§ 35 - 53), Abschnitt 1 - Zustandekommen des Verwaltungsaktes (§§ 35 - 42a) (2017).
- von Lucke, J., & Etscheid, J. (2020). Künstliche Intelligenz im öffentlichen Sektor. *HMD Praxis der Wirtschaftsinformatik*, 57(1), 60–76. <https://doi.org/10.1365/s40702-019-00579-6>
- What is Scopus about?* (2023, September 21). [https://service.elsevier.com/app/answers/detail/a\\_id/15100/supporthub/scopus/related/1/](https://service.elsevier.com/app/answers/detail/a_id/15100/supporthub/scopus/related/1/)

# Anhang

## Anhang A: Interviewleitfaden KI-DemÖ Mitarbeiter

### **Begrüßung:**

Zunächst möchte ich mich bedanken, dass Sie sich Zeit genommen haben für dieses Interview. Damit keine Informationen verloren gehen, würde ich das Interview aufzeichnen wollen. Die Ergebnisse werden anonymisiert verwendet und die Aufzeichnung nach Ende der Arbeit gelöscht. Sind Sie damit einverstanden, dass ich das Gespräch aufzeichne?

### **Vorstellung der Arbeit:**

Für die Bachelorarbeit möchte ich ein Interface entwickeln, das die Ergebnisse eines fiktiven KI-basierten Entscheidungsunterstützungssystems bzgl. der Planung von Kontrollen von Lebensmittelbetrieben nachvollziehbar darstellt.

### **Ziel des Interviews:**

Verständnis dafür bekommen, was die Aufgaben der Lebensmittelkontrolleure sind - vor allem die Arbeit im Innendienst am Computer und da noch konkreter die Planung der Kontrollen. Und wie die dazu gehörenden Abläufe sind.

Wenn Sie bis hierhin keine Fragen haben, würde ich mit meinen Fragen an Sie beginnen.

### **Fragen:**

Ich habe ihre Interviewergebnisse gelesen und werde mich immer mal wieder darauf beziehen.

### **Fragen zur Ihrer Person:**

- Können Sie mir kurz erläutern, in welchem Zusammenhang Sie die Interviews geführt haben?
- Wie viele Interviews haben Sie geführt? Waren Sie bei der Behörde vor Ort und haben den Kontrolleuren bei ihrer Arbeit über die Schulter geschaut?

### **Fragen zu den Aufgaben der Lebensmittelkontrolleure:**

In ihren Interviews wird gesagt: 70% Außendienst/ 30% Innendienst:

- Aus welchen Tätigkeiten besteht der Innendienst? Alles am Computer?
- Aus welchen Aufgaben besteht die Vorbereitung?
- Welche Kennwerte / Parameter sind bei der Vorbereitung der Kontrollen wichtig?
- Welche Bedeutung haben Ergebnisse vorheriger Kontrollen für eine neue Kontrolle?
- Werden vor einer Kontrolle standardmäßig vorherige Kontrollergebnisse rausgesucht und betrachtet?
- Welche Aufgaben /Arbeitsschritte gibt es nach einer Kontrolle?  
Wie werden die Ergebnisse einer Kontrolle weiterverarbeitet?

**Fragen zu der digitalen Infrastruktur ( der Software / den Geräten, die bei der Arbeit eingesetzt werden und den Daten):**

Wie ich u.a. aus den Interviews herausgelesen habe, ist die Software Balvi IP das zentrale Werkzeug im Arbeitsalltag der Lebensmittelkontrolleure.

- Welche der oben genannten Aufgaben sind digitalisiert?
- Welche der oben genannten Aufgaben werden mit Balvi bearbeitet?
- Welche der Aufgaben werden nicht am Computer bearbeitet?
- In welcher Form liegen genutzte Daten vor?  
Textdateien, Tabellen, PDFs
- Kann man sich Balvi als Umsetzung des in Anlage I zu findenden Beispiels...?
- Berücksichtigt Balvi bei der Erstellung der Liste der zu kontrollierenden Betriebe, wie viele Kontrollen ein Kontrolleur schafft?

Im Interview Frage zu „Datenqualität“: „Mäßig – Fehlerhafte Informationen werden häufig erst vor Ort bei der Kontrolle festgestellt:

- Wie sehen fehlerhafte Informationen aus?
- Wie entstehen sie?

Welche Daten- bzw. Informationsquellen, neben den Ergebnissen von Kontrollen und Beschwerden aus der Bevölkerung könnten für die Planung von Kontrollen und dem Ausfindigmachen von Schwarzen Schafen noch relevant sein?

Entscheiden Kontrolleure selbst, ob sie im Außendienst mit Papier oder einem Tablet arbeiten?

### Thema Optimierung /Verwendung von Algorithmen:

„Es ist möglich Kriterien bei der Kontrolle einzugeben, die dazu führen, dass ein Betrieb schneller wieder in der Kontrolle landet (bspw. Unsauberkeit)

- Welche weiteren Kriterien führen noch dazu?

### Fragen zu der Nutzung von Softwaresystemen:

- Ist Ihrer Einschätzung nach das Thema Nachvollziehbarkeit der genutzten Softwaresysteme bzw. die Funktionsweise der Software, also z.B. Balvi, die die Liste der zu kontrollierenden Betriebe erstellt, für die Kontrolleure relevant?
- Sind Mitarbeiter in Behörden dem Einsatz von KI generell offen gegenüber?
- Was glauben Sie ist den Lebensmittelkontrolleuren bei der Nutzung eines Interfaces besonders wichtig?

### AM ENDE:

- Bekommen kontrollierte Betriebe eine Checkliste, der gefundenen Mängel (in Papierform /elektronisch)?
- Was ist daraus geworden: Gibt es die Möglichkeit die Planung der zu kontrollierenden Betriebe mit vorzunehmen? Was war die Antwort auf diese Frage?

### WENN noch etwas Zeit übrig ist:

- Interview Frage „Integrationsgrad der Systeme“:  
Wie ist die Frage gemeint?
- Interview Frage „Stammdatenharmonisierung“ Was ist damit gemeint?  
Warum gab es darauf keine Antwort?

### Schluss:

- Dann wäre ich hiermit am Ende. Danke noch mal, dass Sie sich Zeit genommen haben. Und ich würde dann jetzt die Aufnahme beenden.
- Dürfte ich Sie nochmal kontaktieren, falls ich noch Fragen habe.



## Anhang B: Interviewleitfaden Lebensmittelkontrolleur

### **Begrüßung:**

Danke dass Sie sich die Zeit nehmen. Damit keine Informationen verloren gehen, würde ich gerne eine Audioaufzeichnung des Interviews machen. Die Ergebnisse werden nur anonymisiert verwendet und die Aufzeichnung nach Ende der Arbeit gelöscht. Sind Sie damit einverstanden, dass das Gespräch aufgenommen wird?

### **Vorstellung der Arbeit:**

Das primäre Ziel meiner Bachelorarbeit, ist es, die Ergebnisse eines fiktiven KI-Systems zu erklären. Wenn Menschen auf Basis von KI-Ergebnissen Entscheidungen treffen müssen und der Mensch Spielraum bei seiner Entscheidung hat, muss er nachvollziehen können, wie das KI-System zu ihrem Ergebnis kam. Dafür braucht es ein Interface, dass diese Erklärungen nachvollziehbar liefert.

Das Sekundärziel ist es ein Interface zu gestalten, das die Planung von Lebensmittelkontrollen darstellt. Aber das System soll KI-basiert sein.

### **Ziel des Interviews**

Verständnis dafür bekommen, wie die Abläufe bei der Planung der Kontrollen aussehen (vor allem in Bezug auf Bewertung der bisherigen Kontrollergebnisse von Betrieben und anderen vorhandenen Daten). Wie sind Daten zu den Betrieben organisiert? Welche Aufgaben gibt es bei der Planung der Kontrollen? Welche Herausforderungen ergeben sich bei der Bewertung von Daten?

*Anhand des Zeigens seines Systems, direkt die Fragen stellen und beantworten*

### **Fragen zur Person**

1. Wie sind Sie zum Beruf des Lebensmittelkontrolleurs gekommen?
2. Wie lange schon dabei?
2. Funktion/Rolle in/neben Lebensmittelkontrolle?

### **ABSCHNITT PLANUNG DER KONTROLLEN:**

- Wie läuft die Planung der Kontrollen mit dem aktuellen System, also Balvi, ab? Können Sie mir das anhand des Systems zeigen?
  - Welche Schritte? Wie viele Schritte?
  - Wie sieht die Liste mit den zu kontrollierenden Betrieben aus?
  - Wird bei der Anzeige der „Monatsliste“ angezeigt, worauf man bei den einzelnen Betrieben besonders achten soll?
  - Wie sind Daten zu den Betrieben organisiert?
  - Welche Aufgaben gibt es bei der Planung der Kontrollen?
  - Gibt es Herausforderungen ergeben bei der Bewertung von Daten?
- Wie wichtig sind die Ergebnisse der letzten Kontrollen eines Betriebes bei der Planung der nächsten Kontrollen?
- Wenn Sie die Kontrolle eines Betriebes planen, würden Sie es begrüßen, wenn Balvi IP Ihnen Hinweise gibt, was bei den zu kontrollierenden Betrieben bei den letzten Kontrollen besonders hervorstach und worauf Sie diesmal besonders achten sollten?
- Welche Daten werden zur Zeit verwendet bei der Bewertung von Betrieben?
- Welche Datenquellen sind für Sie noch denkbar?

Haben Sie schon mal gedacht: „Wenn ich diese Information hätte, würde mir das weiterhelfen“?

verschiedene Datenquellen?

- bisherige Kontrollprotokolle
- bisherige Kontrollergebnisse
- Lieferkette
- evtl. Beschwerden aus der Bevölkerung

## **ABSCHNITT ERKLÄRUNGEN / VERSTÄNDLICHKEIT**

- Wenn es mal Probleme mit der Software gibt: Welches Feedback gibt Ihnen die Software bzgl. des Systemstatus?
- Werden Informationen in Balvi in Form von Datenvisualisierungen dargestellt (also Grafiken oder Diagramme)?
- Welche Möglichkeiten bietet ihr aktuelles System es besser zu verstehen?

## **TREFFERQUOTE:**

- Können Sie sich vorstellen, wie man die schwarzen Schafe besser hervorstechen lässt?
- evtl. Rezensionen von Internetbewertungsportalen und Social Media?

Schluss:

- Dann wäre ich hiermit am Ende. Danke noch mal, dass Sie sich Zeit genommen haben. Und ich würde dann jetzt die Aufnahme beenden.
- Dürfte ich Sie nochmal kontaktieren, falls ich noch Fragen habe.

# Nutzerbefragung

## Fragebogen zum A/B-Test des implementierten System

Liebe\_r Teilnehmende, vielen Dank, dass du an dieser Evaluation teilnimmst. Im folgenden Fragebogen geht es um die Nachvollziehbarkeit, die Gebrauchstauglichkeit und das Maß der Belastung des Systems.

### DATENSCHUTZERKLÄRUNG

Bevor wir mit der Umfrage beginnen, benötigen wir Ihr explizites Einverständnis, dass wir Ihre Daten auch speichern und verwenden dürfen. Datenschutz: Die Umfrage wird über die Server der Universität zu Lübeck durchgeführt. Detaillierte Informationen zum Datenschutz an der Universität zu Lübeck gemäß DSGVO finden Sie auf der entsprechenden Datenschutz-Seite. Ihre Angaben werden auf unserem Server gespeichert, ausgewertet und anonymisiert veröffentlicht (z.B. im Rahmen von wissenschaftlichen Publikationen, Konferenzbeiträgen etc.).

☐ Um die Umfrage zu öffnen, akzeptieren Sie bitte unsere Datenschutzerklärung.

Weiter

Made in LimeSurvey ↗

---

## Nach Aufgabe 1 & 2

Sie haben Aufgabe 1 & 2 abgeschlossen. Füllen Sie den ersten Fragebogen aus.

Zurück

Weiter

Made in LimeSurvey ↗

## Beanspruchungshöhe bei Aufgabe 1 und 2

Geben Sie jetzt bitte für jede der unten stehenden Dimensionen an, wie hoch die Beanspruchung war. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

\*  
**Geistige Anforderung**

	Ge- rin g	Ho- ch
Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	

\*  
**Körperliche Anforderungen**

[illegible]

\*  
**Zeitliche Anforderungen**

[illegible]

**\*  
Leistung**

	Gut		Schlecht
Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○

**\*  
Anstrengung**

	Gering		Hoch
Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○

**\*  
Frustration**

	Gering		Hoch
Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○
	○	○	○

Zurück

Weiter

# QUIZ

Wie wird der Risikowert für einen Betrieb berechnet?

Wählen Sie alle zutreffenden Optionen

Bitte wählen Sie eine Antwort

- ☐ Ein KI-System wertet die vorhandenen Daten aus und berechnet das Risiko
- ☐ Ein Experte wertet die vorhandenen Daten aus und entscheidet, wie hoch das Risiko ist
- ☐ Eine Mischung aus beiden

Wie bewertet das System die Datenqualität und wie wird dies dargestellt?

Wählen Sie alle zutreffenden Optionen

- ☐ Die Datenqualität und ihre Unterkategorien werden mit einer Ampelfarbe bewertet, wobei jede Farbe einen bestimmten Qualitätsstatus repräsentiert.
- ☐ Alle Kategorien werden zusammengefasst und als ein einziger numerischer Wert dargestellt.
- ☐ Das System stellt nur die Gesamtqualität der Daten dar, ohne auf einzelne Kategorien einzugehen.

Welche Rolle spielt das System bei der Erklärung der KI-Ergebnisse?

Wählen Sie alle zutreffenden Optionen

- ☐ Es zeigt nur die Endergebnisse ohne Erklärung.
- ☐ Es visualisiert die genauen Gewichtungen jeder Datenquelle im KI-Modell.
- ☐ Es visualisiert Vergleiche mit früheren Fällen.

**Welche Datenquelle wird am höchsten gewichtet?**

*Wählen Sie alle zutreffenden Optionen*

- ☐ Online-Bewertungen
- ☐ Kontrollberichte
- ☐ Selbsteinschätzung der Betriebe

**Inwiefern hilft das System dabei, die Entscheidungen der KI nachzuvollziehen?**

*Wählen Sie alle zutreffenden Optionen*

- ☐ Durch die Darstellung der Entscheidungslogik und des Bewertungsprozesses
- ☐ Indem es ausschließlich das Endergebnis zeigt
- ☐ Durch das Angebot von Hypothesen, die nicht auf realen Daten basieren

**Welche Informationen liefert das System über die Risikofaktoren eines Lebensmittelbetriebs?**

*Wählen Sie alle zutreffenden Optionen*

- ☐ Detaillierte Beschreibungen jedes identifizierten Risikofaktors
- ☐ Nur die Gesamtbewertung des Risikos
- ☐ Keine spezifischen Informationen, nur allgemeine Richtlinien

Zurück

Weiter



## Aufgabe 3

Gehen Sie zurück zum System.

Sie haben Aufgabe 3 abgeschlossen. Füllen Sie jetzt den zweiten Fragebogen aus.

Zurück

Weiter

Made in LimeSurvey 



\*

### Leistung

	Gut											Schlecht
Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

\*

### Anstrengung

	Gering											Hoch
Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

\*

### Frustration

	Gering											Hoch
Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Zurück

Weiter

# SIPA-Fragebogen

Der SIPA-Fragebogen fragt ab, ob das System nachvollziehbar ist.

\*

Bitte geben Sie an, inwieweit Sie den folgenden Aussagen zustimmen bzw. nicht zustimmen.

	stimme über- haupt nicht zu	stimme weitge- hend nicht zu	stimme eher nicht zu	stimme eher zu	stimme weitge- hend zu	stimme völlig zu
Es war für mich transparent, welche Informationen vom System erfasst wurden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Informationen, die das System erfassen konnte, waren für mich einsehbar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es war für mich nachvollziehbar, wie die gesammelten Informationen zu diesem Ergebnis führten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Informationsverarbeitung des Systems war für mich nachvollziehbar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mit den für mich zugänglichen Informationen war das Ergebnis für mich absehbar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Informationsverarbeitung des Systems war für mich vorhersehbar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter

# Bewertung des Systems

\*

Geben Sie bitte Ihre Zustimmung zu den folgenden Aussagen an.

	1 - Stimme gar nicht zu	2	3	4	5 - Stimme voll zu
Ich kann mir sehr gut vorstellen, das System regelmäßig zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde das System als unnötig komplex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde das System als einfach zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass die verschiedenen Funktionen des Systems gut integriert sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich finde, dass es im System zu viele Inkonsistenzen gibt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mir vorstellen, dass die meisten Leute das System schnell zu beherrschen lernen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich empfinde die Bedienung als sehr umständlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Zurück

Weiter

# Demografische Informationen

Wie alt sind Sie?

*In dieses Feld dürfen nur Zahlen eingegeben werden.*

Welches Geschlecht haben Sie?

☐  
weiblich☐  
männlich☒  
Keine Antwort

Was ist Ihr Beruf?

Haben Sie Vorerfahrungen mit KI-Systemen oder mit von einem KI-System erstellten Ergebnissen?

☐  
Ja☐  
Nein☒  
Keine Antwort

Zurück

Absenden

Geschafft. Vielen Dank, dass du dir die Zeit genommen haben und an der Umfrage teilgenommen haben.

# Vielen Dank!

Turn your own questions into answers and start building your own survey today.

[Get started now](#)

Created with  **LimeSurvey**

## Anhang D: Beispieldatensatz

### Auf den folgenden Seiten ist zu sehen:

1. „KI-generierte“ Zusammenfassung eines Kontrollprotokolls
2. „KI-generierte“ Zusammenfassung der Gewerbeanmeldedaten
3. „KI-generierte“ Zusammenfassung von Online-Rezensionen
4. „KI-generierte“ Zusammenfassung der Gesamtauswertung
5. Gewichtungsansicht mit einzelnen Teilen
6. Ansicht der Datenqualität inkl. textlicher Informationen zu den Kategorien
7. Ab hier folgen die „Rohdaten“: 1x Kontrollprotokoll, 1x Gewerbeanmeldedaten, 1x Online-Rezensionen

## 1. Datenquelle Betriebskontrolle 1

---

Das ist eine vom System erstellte Zusammenfassung der Datenquelle. Die Originalquelle befindet sich am Ende dieser Zusammenfassung.

### Kontrollbericht Nr. 1

**Betrieb:** Fischhandel Meeresbrise

**Kontrolldatum:** 22.03.2022

**Kontrollobjekt:** ID-9834-KL

**Kontrollzweck:** planmäßige Routinekontrolle am: 22.03.2022

**Kontrollierte Räume/Bereiche:** Betriebsstätte (allgemein), Küche, Verkaufsraum

**Nicht kontrollierte Räume/ Bereiche:** Personalumkleide, Kühlager

**HACCP-Konzept:** nicht vorhanden

### Feststellungen:

#### Verkaufsraum:

- **Feststellung:** Die Kühlkette war aufgrund eines technischen Defekts im Verkaufsraum für eine unbekannte Dauer unterbrochen. Diese Unterbrechung hatte zur Folge, dass die Temperatur der kühlbedürftigen Produkte nicht durchgehend den gesetzlichen Vorschriften entsprach.

**Anordnung:** Es ist dringend erforderlich, die Kühlgeräte umgehend zu reparieren, um die Qualität und Sicherheit der angebotenen Lebensmittel zu gewährleisten. Verdorbene Produkte sind umgehend zu entsorgen.

- **Feststellung:** Anzeichen von Feuchtigkeit und Schimmelbildung in einigen Bereichen der Kühltheke

**Anordnung:** Umgehende Reinigung und Desinfektion der gesamten Kühltheke



- **Feststellung:** Das HACCP-Konzept des Betriebs ist nicht vollständig umgesetzt. Es fehlen nachweisbare Dokumentationen kritischer Kontrollpunkte, was ein signifikantes Risiko für die Lebensmittelsicherheit darstellt.

**Anordnung:** Die Implementierung eines vollständigen HACCP-Konzepts

**Weitere Informationen:**

- Es wurden Lichtbilder wurden angefertigt, Merkblätter ausgehändigt & Proben entnommen

---

## 2. Datenquelle Gewerbeanmeldedaten

---

Das ist eine vom System erstellte Zusammenfassung der Datenquelle. Die Originalquelle befindet sich am Ende dieser Zusammenfassung.

**Zusammenfassung der Gewerbeanmeldung für „Fischhandel Meeresbrise“**

- **Name des Unternehmens:** Fischhandel Meeresbrise
- **Betriebsinhaber:** Otto Hansen
- **Telefon:** 0431 12345678
- **Handy:** 0176 987654321
- **Betriebsadresse:** Bremer Straße 45, 24143 Kiel

**Betriebsinformationen:**

- **Tätigkeit:** Verkauf von Fisch, Meeresfrüchten, Herstellung von div. Fischspezialitäten, Lieferung von kalten Platten
- **Art des Betriebs:** Lebensmittelhandel, Lieferdienst
- **Beginn:** 01.10.2003
- **Mitarbeiter:** 1 Vollzeit, 1 Teilzeit (ohne Inhaber)

---

## 3. Datenquelle: Online-Bewertungen & -Beschwerden

---

Das ist eine vom System erstellte Zusammenfassung der Datenquelle. Die Originalquelle befindet sich am Ende dieser Zusammenfassung.

**Zusammenfassung der Bewertungen für Fischhandel Meeresbrise:**

- **Schlechte Bewertung #1 (1 von 5 Sternen):** Kritik an mangelnder Hygiene. Beanstandet wurde, dass eine Frau mit Handschuhen Geld anfasste, sich damit in den Mund ging und anschließend ohne Handschuhwechsel das Essen berührte und aushändigte.
- **Schlechte Bewertung #2 (1 von 5 Sternen):** Unzufriedenheit mit dem Krabbensalat. Der Kunde fand das Verhältnis von Majonäse zu Krabben im Salat unangemessen hoch und empfand den Preis als zu teuer, was zu der Entscheidung führte, nicht mehr dort einzukaufen.
- **Mittelmäßige Bewertung #3 (3 von 5 Sternen):** Gemischtes Feedback. Positiv bewertet wurde die Auswahl an Fischen und Meeresfrüchten, insbesondere die Qualität der Garnelen. Jedoch gab es Bedenken bezüglich der Frische des Lachses, dessen Geschmack als akzeptabel, aber nicht herausragend beschrieben wurde.

Es gab keine Rezensionen oder Beschwerden über den Fischhandel Meeresbrise auf Social-Media-Kanälen.

## 4. Zusammenfassung der Auswertung

Betrieb Fischhandel Meeresbrise - Risikobewertung **6.5**

Die Analyse für den "Fischhandel Meeresbrise" integriert verschiedene Datenquellen, um ein umfassendes Bild zu erstellen. Dabei werden die Informationen aus der Gewerbebeanmeldung (20% Gewichtung), den behördlichen Kontrollberichte (70% Gewichtung) und den Onlinebewertungen (10% Gewichtung) berücksichtigt, mit einem besonderen Fokus auf die Hygienestandards.

Bezüglich der Gewerbebeanmeldungsdaten wird deutlich, dass es sich um ein etabliertes Einzelunternehmen handelt, das seit 2003 besteht und durch zwei Mitarbeiter unterstützt wird. Dies lässt auf eine gewisse Erfahrung im Bereich des Lebensmittelhandels schließen.

Die behördlichen Kontrollberichte zeigen jedoch ein gemischtes Bild. Der erste Bericht enthüllte gravierende Mängel, wie eine unterbrochene Kühlkette und Unzulänglichkeiten im HACCP-Konzept. Die darauf folgenden Berichte zeigten zwar Verbesserungen, aber auch die Notwendigkeit weiterer Schulungen und kleinere Mängel im Lagerraum. Diese Entwicklung weist darauf hin, dass der Betrieb auf die Kritik reagiert und sich um Verbesserungen bemüht, aber weiterhin kontinuierliche Aufmerksamkeit für die Aufrechterhaltung und Verbesserung der Hygienestandards benötigt.

Die Onlinebewertungen auf Google zeigen insgesamt eine hohe Kundenzufriedenheit mit einer durchschnittlichen Bewertung von 4,8 aus 150 Bewertungen. Dies deutet auf eine gute allgemeine Wahrnehmung des Betriebs hin. Allerdings sind die negativen Bewertungen, die hygienische Bedenken und Unzufriedenheit mit der Produktqualität zum Ausdruck bringen, besorgniserregend. Diese Bewertungen, insbesondere die Schilderung hygienischer Mängel, stehen im Widerspruch zu den insgesamt positiven Kontrollberichten und bedürfen einer genaueren Untersuchung und Reaktion seitens des Betriebs.

Die Zusammenfassung der Daten zeigt, dass "Fischhandel Meeresbrise" in einigen Bereichen gut abschneidet, insbesondere im Hinblick auf die Kundenzufriedenheit und die Reaktion auf frühere Kritikpunkte. Dennoch heben die Onlinebewertungen und die fortgesetzten kleineren Mängel in den Kontrollberichten die Notwendigkeit hervor, die Hygienestandards konsequent einzuhalten und das Personal regelmäßig zu schulen. Insgesamt kann dem Betrieb eine Note "befriedigend bis gut" gegeben werden.

Die Analyse für den "Fischhandel Meeresbrise" integriert verschiedene Datenquellen, um ein umfassendes Bild zu erstellen. Dabei werden die Informationen aus der Gewerbebeanmeldung (20% Gewichtung), den behördlichen Kontrollberichten (70% Gewichtung) und den Onlinebewertungen (10% Gewichtung) berücksichtigt, mit einem besonderen Fokus auf die Hygienestandards.

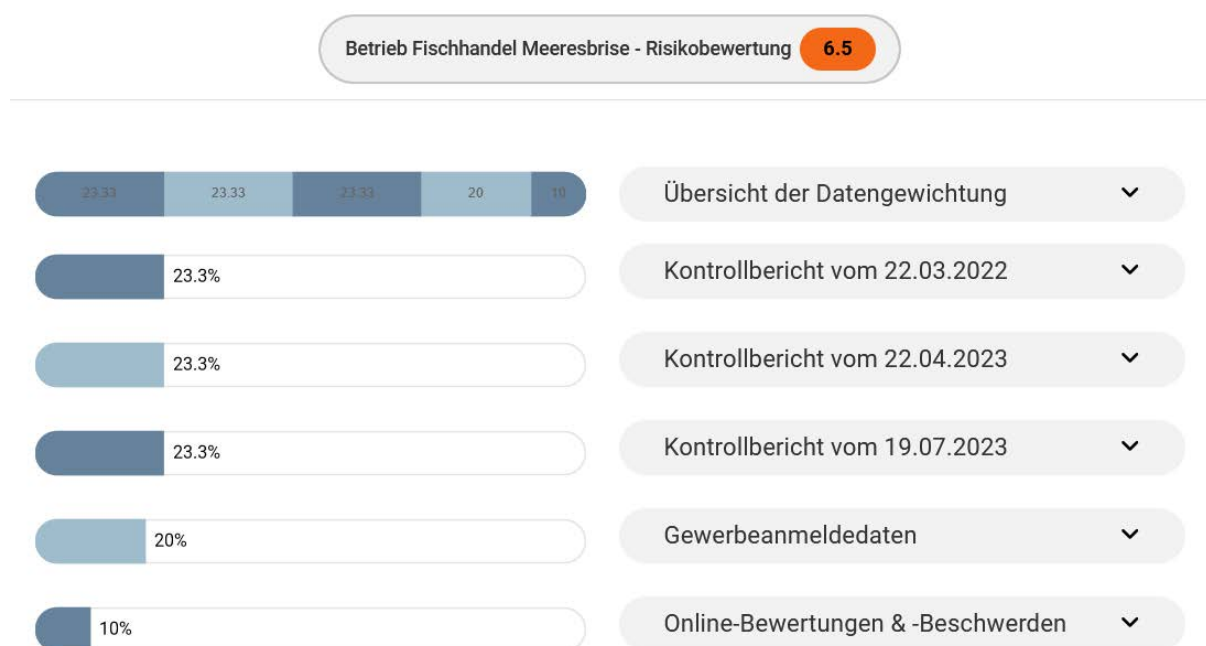
Bezüglich der Gewerbebeanmeldungsdaten wird deutlich, dass es sich um ein etabliertes Einzelunternehmen handelt, das seit 2003 besteht und durch zwei Mitarbeiter unterstützt wird. Dies lässt auf eine gewisse Erfahrung im Bereich des Lebensmittelhandels schließen.

Die behördlichen Kontrollberichte zeigen jedoch ein gemischtes Bild. Der erste Bericht enthüllte gravierende Mängel, wie eine unterbrochene Kühlkette und Unzulänglichkeiten im HACCP-Konzept. Die darauf folgenden Berichte zeigten zwar Verbesserungen, aber auch die Notwendigkeit weiterer Schulungen und kleinere Mängel im Lagerraum. Diese Entwicklung weist darauf hin, dass der Betrieb auf die Kritik reagiert und sich um Verbesserungen bemüht, aber weiterhin kontinuierliche Aufmerksamkeit für die Aufrechterhaltung und Verbesserung der Hygienestandards benötigt.

Die Onlinebewertungen auf Google zeigen insgesamt eine hohe Kundenzufriedenheit mit einer durchschnittlichen Bewertung von 4,8 aus 150 Bewertungen. Dies deutet auf eine gute allgemeine Wahrnehmung des Betriebs hin. Allerdings sind die negativen Bewertungen, die hygienische Bedenken und Unzufriedenheit mit der Produktqualität zum Ausdruck bringen, besorgniserregend. Diese Bewertungen, insbesondere die Schilderung hygienischer Mängel, stehen im Widerspruch zu den insgesamt positiven Kontrollberichten und bedürfen einer genaueren Untersuchung und Reaktion seitens des Betriebs.

Die Zusammenfassung der Daten zeigt, dass "Fischhandel Meeresbrise" in einigen Bereichen gut abschneidet, insbesondere im Hinblick auf die Kundenzufriedenheit und die Reaktion auf frühere Kritikpunkte. Dennoch heben die Onlinebewertungen und die fortgesetzten kleineren Mängel in den Kontrollberichten die Notwendigkeit hervor, die Hygienestandards konsequent einzuhalten und das Personal regelmäßig zu schulen. Insgesamt kann dem Betrieb eine Note "befriedigend bis gut" gegeben werden.

## 5. Gewichtungsansicht



## Übersicht der Datengewichtung

Die Bewertung aller zur Verfügung stehenden Datenquelle durch das KI-System ergab eine Risikobewertung von **6.5**

Die Untersuchung zeigte deutlich, dass die fortwährenden und sorgfältigen Kontrollmaßnahmen maßgeblich zur Steigerung der hygienischen Standards im Unternehmen beigetragen haben. Es wurde ein klarer, positiver Zusammenhang zwischen der Intensivierung der Überwachungsprozesse und der signifikanten Verbesserung der hygienischen Bedingungen innerhalb des Betriebs festgestellt.

Für die Risikobewertung des Betriebs "Fischhandel Meeresbriese" konnten folgende Datenquellen herangezogen werden:

- 3 Betriebskontrollen in den letzten 12 Monaten
- mehrere Rezensionen von Bewertungsportalen und Social Media
- Gewerbeanmeldedaten

## Kontrollbericht vom 22.03.2022

Die Bewertung der Datenquelle "Kontrollbericht vom 22.03.2022" basiert auf der Auswertung der KI.

Die Risikobewertung für die spezifische Datenquelle "Kontrollbericht vom 22.03.2022" ergibt einen Wert von **9.1**

Allgemeine Gewichtung der Datenquelle: sehr hoch

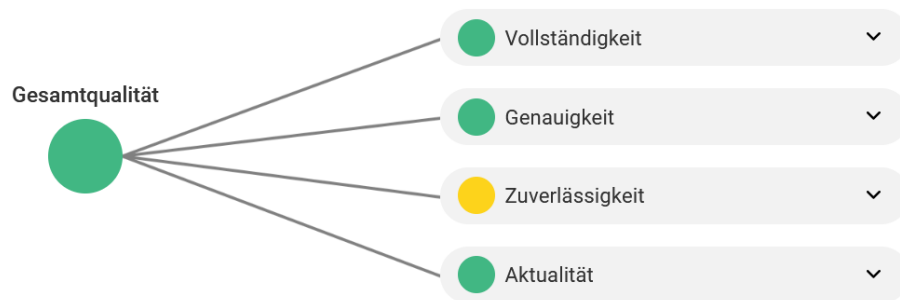
Kontrollzweck: planmäßige Routinekontrolle am 22.03.2022.

Identifizierte Themen und Informationen:

- Verkaufsraum & Kühltheke:
- Technischer Defekt in der Kühltette: Unterbrechung der Kühltette im Verkaufsraum für unbekannte Dauer
- Anordnung der Reparatur der Kühlgeräte und Entsorgung der betroffenen Lebensmittel.
- Anzeichen für Feuchtigkeit und Schimmel in einigen Bereichen der Kühltheke
- Umgehende Reinigung und Desinfektion der gesamten Kühltheke wurde angeordnet
- Mangelhafte Umsetzung des HACCP-Konzepts: Fehlende nachweisbare Dokumentationen kritischer Kontrollpunkte.
- Anordnung zur Implementierung eines vollständigen HACCP-Konzepts

[zur Datenquelle](#)

## 6. Ansicht der Datenqualität



### Vollständigkeit

Nach sorgfältiger Überprüfung aller relevanten Informationen und Quellen lässt sich feststellen, dass die Datenlage als umfassend und vollständig betrachtet werden kann, was eine solide Grundlage für weitere Analysen bietet.

### Genauigkeit

Die Analyse der Kundenfeedbacks hat eine Übereinstimmung aufgezeigt: Einige wenige negative Kundenbewertungen spiegeln sich konsistent in den Ergebnissen der Kontrollberichte wider. Dies deutet auf eine hohe Genauigkeit und Verlässlichkeit der Daten hin.

### Zuverlässigkeit

Bei der Beurteilung der Zuverlässigkeit der Datenquellen „Onlinebewertungen“ und „Beschwerden aus der Bevölkerung“ bestehen gewisse Unsicherheiten. Diese Quellen können subjektiven Verzerrungen unterliegen, was die Konsistenz und Verlässlichkeit der Daten potenziell beeinträchtigt.

### Aktualität

Bei der Überprüfung der Datenbestände zeigt sich, dass sie auf dem neuesten Stand sind. Die Informationen wurden regelmäßig und zeitnah aktualisiert, was sicherstellt, dass die Daten aktuell und relevant sind.

# Kontrollbericht Nr. 1

Bericht zur Kontrolle gemäß Artikel 13 der VO (EU) Nr. 2017/625

Kontrolldatum: 22.03.2022

Kontrollobjekt: ID-9834-KL

## Betrieb:

Fischhandel Meeresbrise

## Kontrollierende Behörde:

### Lebensmittelkontrolle Kiel

Hauptstraße 14, 12345 Kiel

Tel.: 0451 1234567

Fax: 0451 7654321

Email: info@lebensmittelkontrolle-kiel.de

## Kontrollzweck:

planmäßige Routinekontrolle am: 22.03.2022

## Betriebsräume-/bereiche:

Kontrolliert: Betriebsstätte (allgemein), Küche, Verkaufsraum

Nicht kontrolliert: Personalumkleide, Kühllager

## Kontrollpunkte:

- ☐ Hygiene-betriebl. Eigenkontrolle
  - ☐ HACCP
  - ☐ Rückverfolgbarkeit
  - ☐ Reinigung/Desinfektion
  - ☐ Ausbildung/Schulung Personal
  - ☐ Schädlingsbekämpfung
  - ☐ Dokumentation
- ☐ Hygiene allgemein
- ☐ Zusammensetzung- nicht mikrob.
- ☐ Kennzeichnung/Aufmachung
- ☐ Andere Kontrollpunkte

**Folgende Bereiche waren in Ordnung: -**

**Verkaufsraum:**

- **Feststellung:** Die Kühlkette war aufgrund eines technischen Defekts im Verkaufsraum für eine unbekannte Dauer unterbrochen. Diese Unterbrechung hatte zur Folge, dass die Temperatur der kühlbedürftigen Produkte nicht durchgehend den gesetzlichen Vorschriften entsprach.  
**Anordnung:** Es ist dringend erforderlich, die Kühlgeräte umgehend zu reparieren, um die Qualität und Sicherheit der angebotenen Lebensmittel zu gewährleisten. Verdorbene Produkte sind umgehend zu entsorgen.  
**Norm:** Art. 4 Abs. 2 VO (EG) Nr. 852/2004 i.V.m. Anh. II Kap. 1 Nr. 1
- **Feststellung:** Anzeichen von Feuchtigkeit und Schimmelbildung in einigen Bereichen der Kühltheke  
**Anordnung:** Umgehende Reinigung und Desinfektion der gesamten Kühltheke  
**Norm:** Art. 4 Abs. 2 VQ (EG) Nr. 852/2004 i.V.m. Anh. II Kap. 1 Nr. 1
- **Feststellung:** Das HACCP-Konzept des Betriebs ist nicht vollständig umgesetzt. Es fehlen nachweisbare Dokumentationen kritischer Kontrollpunkte, was ein signifikantes Risiko für die Lebensmittelsicherheit darstellt.  
**Anordnung:** Die Implementierung eines vollständigen HACCP-Konzepts ist unerlässlich, um eine kontinuierliche Einhaltung der lebensmittelhygienischen Anforderungen sicherzustellen und potenzielle Gefahren für die Verbraucher zu minimieren.  
**Norm:** Art. 4 Abs. 2 VQ (EG) Nr. 852/2004 i.V.m. Anh. II Kap. 1 Nr. 1

**Weitere Informationen:**

- **Lichtbilder wurden angefertigt** ☒
- **Merkblätter wurden ausgehändigt** ☒
- **Probe wurden entnommen** ☒
- **Diese Kontrolle ist gebührenpflichtig** ☐

**Anordnung der sofortigen Vollziehung:**

Für die Anordnungen ordne ich die sofortige Vollziehung an. Aufgrund der festgestellten Mängel in Ihrem Betrieb, insbesondere bezüglich der unzureichend gekühlten Produkte und fehlenden Dokumentation, besteht ein Risiko für die Gesundheit der Verbraucher. Die sofortige Vollziehung der genannten Anordnungen dient dem Schutz der Öffentlichkeit. Ihre wirtschaftlichen Interessen müssen in diesem Fall zurückstehen. Ein Widerspruch hätte keine aufschiebende Wirkung.

**Rechtsbehelfsbelehrung:**

Soweit in diesem Kontrollbericht etwaige Anordnungen getroffen worden sind, haben Sie das Recht, innerhalb eines Monats nach Erhalt des Kontrollberichts bei Stadt Kiel Widerspruch zu erheben.

Gegen die Anordnung der sofortigen Vollziehung kann beim Schleswig-Holsteinischen Verwaltungsgericht in Schleswig, Gerichtsallee 5, 24837 schriftlich, elektronisch oder zur Niederschrift des Urkundsbeamten der Geschäftsstelle, ein Antrag auf Wiederherstellung der aufschiebenden Wirkung gem. §80 Abs. 5 der Verwaltungsgerichtsordnung gestellt werden.

**Erklärung des/der Verantwortlichen (ggf. der anwesenden Person):**

**Verantwortliche(r)/ Beauftragte(r):** Stefan Müller **Unterschrift:**\_\_\_\_\_



# Daten aus der Gewerbeanmeldung für Fischhandel Meeresbrise

## Unternehmen:

**Name:** Fischhandel Meeresbrise

**Rechtsform:** Einzelunternehmen

## Betriebsinhaber:

**Name:** Otto Hansen

**Anschrift:** Musterstraße 23 45, 24113 Kiel

**Telefon:** 0431 12345678

**Handy:** 0176 987654321

**Email:** -

## Betriebsadresse:

**Name:** Fischhandel Meeresbrise

**Anschrift:** Bremer Straße 45, 24143 Kiel

## Betriebsinformationen:

**Tätigkeit:** Verkauf von Fisch, Meeresfrüchten, Herstellung von div. Fischspezialitäten,  
Lieferung von kalten Platten

**Art des Betriebs:** Lebensmittelhandel, Liefersdienst

**Status:** Haupterwerb

**Beginn:** 01.10.2003

**Mitarbeiter:** 1 Vollzeit, 1 Teilzeit (ohne Inhaber)

**Früherer Gewerbetreibender oder Firmenname:** -

**Erlaubnis:** Ja, von Behörde Kiel

## **Bewertung auf Bewertungsportalen, wie "Google" oder "Yelp"**

### **Schlechte Bewertung für Fischhandel Meeresbriese #1:**

★☆☆☆☆ (1 von 5 Sternen)

"Sehr unhygienisch! Blonde Frau hat blaue Handschuhe an und fässt damit Geld an, geht sich damit in den Mund und fässt dann das Essen an und gibt es aus. Einfach nur widerwärtig!"

### **Schlechte Bewertung für Fischhandel Meeresbriese #2:**

★☆☆☆☆ (1 von 5 Sternen)

"Ich habe heute einen Krabbensalat bei Ihnen gekauft...Ca. 300 Gramm für 12,16 €...Sie sollten den Salat anders bezeichnen, da der Majonäsenanteil im Verhältnis zu den Krabben viel zu hoch ist. Für mich war das der letzte Einkauf bei Ihnen."

### **Mittelmäßige Bewertung für Fischhandel Meeresbriese #3:**

★★★☆☆ (3 von 5 Sternen)

"Die Auswahl an Fischen und Meeresfrüchten war ansehnlich. Ich habe frische Garnelen und Lachs gekauft. Die Garnelen waren von guter Qualität, frisch und sauber. Beim Lachs hingegen war ich nicht ganz überzeugt. Er sah frisch aus, aber beim Zubereiten zu Hause bemerkte ich, dass der Fisch doch nicht mehr der frischeste war. Der Geschmack war in Ordnung, aber nicht herausragend."

## **Beschwerden über ein Produkt/Betrieb auf social-media Kanälen**

Es wurden keine Rezensionen oder Beschwerden auf Social Media Kanälen gefunden

# Erklärung

Ich versichere an Eides statt, die vorliegende Arbeit selbstständig verfasst und nur die angegebenen Quellen benutzt zu haben.

---

Kiel, den 03.12.2023, Ilja Jamkis