Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Konzeption und Evaluation eines Modells zur Unterstützung des Designs von Erklärungen in erklärbaren Systemen

Conception and evaluation of a model to support the design of explanations in explainable systems

Masterarbeit

im Studiengang Informatik

von

Florian Martin Herzog

Prüfer: Prof. Dr. Kurt Schneider
Zweitprüferin: Dr. Jil Ann-Christin Klünder
Betreuerin: M. Sc. Larissa Chazette

Hannover, 15.10.2021

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 15.10.2021

Florian Martin Herzog

Zusammenfassung

Konzeption und Evaluation eines Modells zur Unterstützung des Designs von Erklärungen in erklärbaren Systemen

Die zunehmende Komplexität von Softwaresystemen und ihre tiefgreifende Integration in das tägliche Leben der Nutzer wecken den Bedarf an transparenter, nachvollziehbarer und vertrauenswürdiger Software. Frühere Arbeiten haben bereits gezeigt, dass Erklärbarkeit als nicht-funktionale Anforderung (NFR) einen signifikanten Einfluss auf diese Qualitätsaspekte sowie auf die Gesamtqualität von Softwaresystemen hat.

Da Erklärbarkeit jedoch eine relativ neue NFR ist, gibt es noch keine Artefakte wie Richtlinien oder Modelle, die Fachleute bei der Identifizierung und Operationalisierung von Anforderungen im Zusammenhang mit Erklärbarkeit unterstützen. Daher ist es wichtig, dass diese Artefakte entwickelt werden, um den Requirements-Engineering-Prozess für Erklärbarkeit und seine Umsetzung zu erleichtern.

Ein Leitfaden über die verschiedenen Möglichkeiten zur Integration von Erklärungen in Softwaresysteme kann die Gestaltung von Erklärungen in erklärbaren Systemen unterstützen. Daher wurde eine Literaturrecherche durchgeführt, um die externen Abhängigkeiten, Eigenschaften und Bewertungsmethoden von Erklärungen zu identifizieren. Ein exemplarischer Einsatz in der Praxis diente schließlich dazu, die Anwendbarkeit eines entwickelten Leitfadens zu evaluieren. Dies hat zu vielversprechenden Ergebnissen hinsichtlich der evaluierten Qualität der resultierenden Erklärungen und des Nutzens des Leitfadens während der Entwicklung von Erklärungen geführt.

In dieser Arbeit wird zusammenfassend ein Leitfaden vorgestellt, der Vorschläge für die Entwicklung von Erklärungen, einen Katalog von Zusammenhängen zwischen Merkmalen von Erklärungen und Softwarequalitätsaspekten sowie Heuristiken für die Erklärungsgestaltung enthält. In zukünftigen Arbeiten sollte der Leitfaden in weiteren Kontexten evaluiert und verbesserte Versionen der abgeleiteten Erklärungen auf Basis der Evaluationsergebnisse entwickelt werden.

Abstract

Conception and evaluation of a model to support the design of explanations in explainable systems

The growing complexity of software systems and their profound integration into users' daily lives, awakens the need for transparent, traceable, and trustworthy software. Previous work has already shown a significant impact of explainability, as a non-functional requirement (NFR), on these quality attributes as well as the overall quality of software systems.

However, because explainability is a relatively new NFR, artifacts such as guidelines or models do not yet exist to assist professionals in identifying and operationalizing requirements related to explainability. Therefore, it is important that these artifacts are in place to facilitate the requirements engineering process for explainability and its implementation.

A guideline with various possibilities for the integration of explanations into software systems may support the design of explanations in explainable systems. For this purpose, I conducted a literature review to identify the external dependencies, characteristics, and evaluation methods of explanations. Finally, an exemplary use in practice served to evaluate the applicability of a developed guideline, which lead to promising results concerning the evaluated quality of the resulting explanations and usefulness of the guideline concerning explanation development.

This thesis presents a guideline containing proposals for the development of explanations, together with a catalog of correlations between the characteristics of explanations and software quality aspects, as well as heuristics for explanation design. As a future contribution, the guidelines have to be evaluated in more contexts and improved versions of the derived explanations based on the evaluation results should be developed.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Einleitung

1.1 Motivation

Die Integration von Softwaresystemen in den Alltag der meisten Menschen wird heutzutage immer tiefer und umfassender [1]. Gleichzeitig steigt die Größe und Komplexität dieser Systeme. Als Lösung für eine wachsende Zahl an Aufgaben reicht die Nutzung von Software dabei von der Routenberechnung in Navigationssystemen, über das Vorschlagen von Produkten bis zu klinischen Unterstützungssystemen [2, 3, 4]. Daher basieren immer mehr getroffene Entscheidungen auf Algorithmen oder den zugrunde liegenden Daten. Vermehrt werden dabei vorwiegend für den Endnutzer schwer nachvollziehbare Algorithmen beispielsweise aus dem Bereich des maschinellen Lernens (ML) verwendet. Nutzer stoßen so öfter auf das Problem, dass die eigene Annahme über die Funktionsweise eines Systems nicht mit der wirklichen System-Funktionsweise übereinstimmt oder unvollständig ist [5].

Folglich wächst der Bedarf an Software, die transparent, nachvollziehbar und vertrauenswürdig ist, um nicht nur das Verständnis der Nutzer für die Funktionsweise des Systems zu erhöhen, sondern auch Diskriminierung durch Software vorzubeugen. Neben bereits bekannten Anforderungen an die Softwarequalität [6] entsteht demnach ein zunehmender Bedarf an Erklärungen innerhalb von Softwaresystemen [7].

Bereits heute existieren Bereiche, in denen für Nutzer ein „Recht auf Erklärung“ gesetzlich verankert ist. Auf europäischer Ebene regelt die *Datenschutz-Grundverordnung* (DSGVO) [8] im Rahmen der Transparenzverpflichtung ein Recht auf die Aufklärung über die Art der Verarbeitung von personenbezogenen Daten und die Auswirkungen einer automatisierten, daraus resultierenden Entscheidung. Eine vergleichbare Verordnung für das Finanzwesen findet sich auch im amerikanischen Recht [9]. So soll sichergestellt werden, dass Unterstützungssysteme, wie sie beispielsweise bei der Bewertung der Kreditwürdigkeit von Personen oder in der Versicherungsbranche angewendet werden, objektive und diskriminierungsfreie Empfehlungen geben. Unter die Verarbeitung von personenbezogenen Daten fallen aber auch die Sortierungs- und Empfehlungsalgorithmen sozialer Netzwerke. Erste Umsetzungen solcher Erklärungen zeigen Instagram[1] für die Priorisierung von

[1] about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works, besucht: 01.10.21

Beiträgen und Tiktok[2] für die Empfehlung von Videos [10, 11]. Instagram erklärt beispielsweise, welche Daten in den Algorithmus einfließen, der die Reihenfolge der Beiträge im Feed von Nutzern festlegt. Hier wird von Instagram vorrangig im Rahmen des Datenschutzes erläutert, welche Daten nicht verwendet und wie verwendete Daten geschützt werden.

Aus dem wachsenden Bedarf und Einsatz von Erklärungen entsteht auch die Notwendigkeit einer Formalisierung der Betrachtung dieser. Frühe Ansätze für erklärbare, intelligente Systeme (*Explainable Artificial Intelligence*: XAI) stammen bereits von Byrne und Cawsey, welche die Erklärungsinhalte [12] und die Interaktion mit Erklärungen untersuchen [13]. Aktuelle Studien auf dem Gebiet befassen sich i. d. R. entweder mit dem technischen Verständnis komplexer Algorithmen zum Beispiel aus dem Bereich des maschinellen Lernens [14, 15, 16] oder dem Einfluss von Erklärungen auf die vom Nutzer wahrgenommene Qualität (externe Qualität [6]) [17, 18, 7]. Letztere wird in dieser Arbeit näher betrachtet.

Bei der Betrachtung von Erklärungen wird Erklärbarkeit (*Explainability*) als neue Nicht-Funktionale Anforderung (*Non Functional Requirement* (NFR)) interpretiert, um Abhängigkeiten mit anderen Qualitätsanforderungen zu untersuchen [19, 2]. Erste Ergebnisse auf diesem Gebiet zeigen, dass Erklärbarkeit eine starke Auswirkung auf die Gesamtqualität von Softwaresystemen hat. Auch kann bereits gesagt werden, dass die Integration von Erklärungen in der Regel mit Kompromissen einhergeht, da *Explainability* sowohl positive als auch negative Auswirkungen auf andere Qualitätsaspekte haben kann. Folglich müssen Qualitätsziele gegeneinander abgewogen werden.

Bis dato existieren vorwiegend Studien, die die Auswirkung einzelner Eigenschaften von Erklärungen auf bestimmte Qualitätsaspekte untersuchen oder einen Überblick über die verschiedenen Einsätze von Erklärungen innerhalb eines bestimmten Kontextes (z. B. Empfehlungssysteme [17]) geben.

Da es sich bei *Erklärbarkeit* um eine neue NFR handelt, fehlen zum aktuellen Zeitpunkt Artefakte, die bei der Anforderungserhebung (Requirements Engineering) für Erklärungen und dessen Operationalisierung unterstützen. Um die Umsetzung im Requirements-Engineering-Prozess zu erleichtern, werden Richtlinien und Modelle benötigt, welche einen Überblick über Vorgehensweisen zur Entwicklung und über Aspekte von Erklärungen geben.

1.2 Lösungsansatz

Auf der Basis der Definitionen für *Erklärbarkeit* von Chazette, Brunotte und Speith wird in dieser Arbeit ein Leitfaden vorgestellt, welcher einen Überblick über die Aspekte von Erklärungen in Softwaresystemen gibt. In einer Literaturrecherche wurden dabei (i) die äußeren Einflüsse auf Erklärungen, (ii) der Bedarf und die Granularität von Erklärungen, sowie (iii) die Evaluation von Erklärungen herausgearbeitet. Diese sind in einem Modell für Erklärungen gebündelt. Auf der Basis werden außerdem die in vorangegangenen Arbeiten gezeigten Zusammenhänge zwischen den einzelnen Aspekten von Erklärungen, sowie dessen Auswirkungen auf

[2] https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you/, besucht: 18.08.2021

die Softwarequalität zusammengefasst.

Dabei konzentriert sich diese Arbeit auf die Qualitätsaspekte, anhand welcher die Qualität von Erklärungen in vorangegangenen Arbeiten bereits indirekt gemessen wurde. Dies hilft unter anderem dabei, die Kompromisse bei der Integration von Erklärungen abzuschätzen. Abschließend werden Design-Empfehlungen für Erklärungen abgeleitet. Die Ergebnisse aus dem Modell, den Zusammenhängen der einzelnen Aspekte sowie den Design-Empfehlungen werden in einem Leitfaden zusammengefasst.

Im zweiten Teil der Arbeit werden die Resultate hinsichtlich der Anwendbarkeit in der Wirtschaft untersucht. Im Zuge dessen werden Erklärungen anhand des in dieser Arbeit entwickelten Leitfadens in ein bestehendes Navigationssystem integriert und evaluiert. Zunächst wurden auf Grundlage der Ergebnisse einer Nutzer-Feedback-Analyse in einem Workshop Probleme identifiziert, welche durch die Integration von Erklärungen gelöst werden sollen. Darauf aufbauend sind Ziele und Evaluationsmöglichkeiten der Erklärungen sowie Ideen zur Umsetzung zusammengetragen worden. Anhand dessen sind konkrete Anforderungen an Erklärungen formuliert, und in dem Navigationssystem umgesetzt worden. Abschließend wurden die integrierten Erklärungen in einer Case-Study sowie einem Quasi-Experiment evaluiert.

1.3 Struktur der Arbeit

Zusammenfassend bietet diese Arbeit einen Leitfaden, welcher die Integration von Erklärungen in bestehende Software erleichtert und als Einstiegspunkt für die Entwicklung detaillierter Modelle für verschiedene Kontexte gesehen werden kann. Dieser Leitfaden unterstützt mit dem enthaltenen Modell die Anforderungserhebung für Erklärungen. Zudem geben die existierenden Ergebnisse zur Integration Hilfestellung bei der Umsetzung von Erklärungen sowie deren Evaluation.

Im nächsten Kapitel (Kapitel 2) werden die Grundlagen von Erklärbarkeit und NFRs im Allgemeinen sowie verwandte Arbeiten vorgestellt. Kapitel 3 definiert das Forschungsziel und beschreibt die verwendeten Methoden. Die Anwendung dessen folgt in den Kapiteln 4 - 6. Wobei Kapitel 4 die durchgeführte Literaturrecherche, Kapitel 5 den resultierenden Leitfaden bestehend aus Modell und gezeigten Zusammenhängen und Kapitel 6 die Evaluation des Leitfadens mithilfe von Nutzern darstellt. Abschließend werden alle Ergebnisse in den Kapitel 7 und 8 diskutiert, auf ihre Allgemeingültigkeit hin untersucht und zusammengefasst.

Kapitel 2

Grundlagen und Verwandte Arbeiten

In diesem Kapitel werden verwandte Arbeiten und die Grundlagen von Erklärbarkeit sowie deren Zusammenwirkung mit ausgewählten Qualitätsaspekten beschrieben.

2.1 Erklärungen in erklärbaren Systemen

Erklärungen in erklärbaren Systemen werden unter dem Oberbegriff Erklärbarkeit (*Explainability*) erforscht. Dabei wird das Integrieren dieser in Softwaresysteme und die daraus resultierenden Auswirkungen auf die Softwarequalität untersucht. *Explainability* wird von unterschiedlichen Autoren durch verschiedene Synonyme beschrieben. Brennen hat einen Katalog zusammengestellt, welcher einige Synonyme zusammenfasst [20]. Dieser enthält unter anderem die Begriffe *Transparency*, *Understandability* und *Interpretability*. In der jüngeren Vergangenheit haben verschiedene Autoren diese Begriffe zum Teil mit unterschiedlichen Definitionen in einen Zusammenhang mit *Explainability* gestellt [7, 5, 19, 21].

Transparency wird als der Grad, zu dem ein System Einblick in dessen Funktionsweise gewährt, beschrieben [7]. Diese Offenlegung kann dabei verschiedene Aspekte von Systemen wie zugrundeliegende Algorithmen z. B. in Empfehlungssystemen [22] oder trainierte Modelle des Maschinellen Lernens [23] betreffen.

Das Verstehen von Erklärungen wird unter *Understandability* zusammengefasst [24]. Das so erlangte Verständnis von Systemnutzern ist folglich als subjektiver Einflussfaktor für Erklärungen zu werten [7]. Für diesen subjektiven Faktor nutzen Wang sowie Balog und Radlinski den Begriff *Perceived Transparency* als Synonym, um die subjektive Aufnahme von *Transparency* bei verschiedenem Verständnis durch Nutzer zu verdeutlichen [21, 22].

Während *Interpretability* zum Teil mit *Understandability* gleichgesetzt wird [7], nimmt die Verwendung des Begriffs für den Grad der möglichen Interpretation der Ausgaben von ML-Algorithmen zu [25]. Als Abgrenzung zwischen *Explainability* und *Interpretability* wird ersteres dabei auch als *Top-Down*-Verständnis und letzteres als *Bottom-up*-Verständnis beschrieben [26]. Die Benutzer folglich können ein System sowohl auf einer niedrigen Ebene interpretieren (*Top-Down*) als auch auf einer hohen Ebene die Funktion eines Systems verstehen (*Bottom-up*).

Da der Fokus dieser Arbeit auf externen Qualitätsaspekten liegt, ist die *Interpretability* von ML-Algorithmen kein Bestandteil dieser Arbeit. Um trotz dessen einer Irritation durch doppelt belegte Begriffe vorzubeugen, wird für den subjektiven Faktor des Verständnisses von Softwaresystem *Perceived Transparency* verwendet. Der Begriff schließt im Kontext dieser Arbeite auch *Understandability* ein.

Um Erklärungen an sich zu definieren, gibt es verschiedene Ansätze. Einig sind sich viele Autoren, dass Erklärungen beim Verständnis von Systemen helfen können. Das heißt, Erklärungen sind ein Lösungsansatz, um das *Mental Model* der *Nutzer* mit der wirklichen Funktionsweise des Systems in Einklang zu bringen. Das *Mental Model* repräsentiert dabei die Annahmen, die Nutzer über die Funktionsweise eines Systems haben. [27]. Norman definiert diesen Unterschied zwischen dem *Mental Model* und der wirklichen Funktionsweise als „Gulf of Evaluation“[28]. Genauer beschreibt er, dass Nutzer in einer solchen Situation die Ausgaben von Softwaresystemen nicht richtig wahrnehmen oder gewünschte Aktionen nicht finden können. Die Nutzer verstehen also das System nicht richtig (*Perceived Transparency*).

Ein Ansatz, Erklärungen für die Schließung von Verständnislücken zu definieren, ist, Erklärungen als Sequenz von Informationen aufzufassen [vgl. 21]. Sovrano, Vitali und Palmirani beschreiben bei diesem Ansatz, dass ein Informationskorpus zum Erhöhen von Verständnis für erklärbare Daten oder Prozesse zur Zufriedenheit der Nutzer eingesetzt wird [übersetzt vgl. 23]. Außerdem fügen sie hinzu, dass die Adressaten der Erklärung mit dem erklärten Systemteil (*explanadum*) zur Erfüllung bestimmter Ziele in einem bestimmten Kontext interagieren. Unterstützt wird dies durch Zahedi et al., die Erklärungen als „Aktualisierung des menschlichen *Mental Model*“ definieren [übersetzt vgl. 29]. Zusätzlich formulieren die Autoren die folgenden drei Bedingungen an Erklärungen:

1. Das aktualisierte *Mental Model* soll nach dem Geben einer Erklärung in der Lage sein, dass die Nutzer ihre Aufgabe bestmöglich durchführen können.

2. Die Nutzer sollen in der Lage sein, den System-Status, die Ziele des Systems und die Bedingungen und Effekte von Aktionen des Systems zu kennen.

3. Eine Erklärung soll der minimal nötigen Aktualisierung des *Mental Model* entsprechen, welche 1. und 2. erfüllt.

Welche Informationen benötigt werden, um die obigen Bedingungen zu erfüllen, wurde in vorangegangenen Arbeiten bereits untersucht. Um verschiedene Inhaltstypen voneinander abzugrenzen, wurde der Inhalt zum Teil als Antwort auf verschiedene Fragewörter definiert (*why*, *what*, *how*) [30]. Am häufigsten tritt dabei die Frage auf, **warum** ein System sich auf bestimmte Weise verhält [2]. Folglich interessieren sich Nutzer vor allem für kausale Zusammenhänge. Problematisch ist hierbei die Einordnung der Inhalte mittels Fragewörtern, da es möglich ist mit verschiedenen Fragewörtern den gleichen Inhalt zu erfragen. Daher diskutiert Wang, dass ein anderes Schema zur Einordnung nötig ist [21].

Eine große Herausforderung von Erklärungen ist allerdings, dass Nutzer einen unterschiedlichen Bedarf für den Inhalt bzw. die transportierten Informationen von Erklärungen haben [2]. Folglich können die Bedingungen 1. und 2. nur

schwer gleichzeitig erfüllt werden. Als Lösungsansatz dafür wird unter anderem das interaktive Anfordern von Erklärungen durch die Nutzer eines Systems genannt [7].

Da die transportierten Informationen eine Kerneigenschaft von Erklärungen darstellen, werden in dieser Arbeit Erklärungen als Information zum Schließen des *Gulf of Evaluation* [28] gesehen. Das bedeutet das *Mental Model* wird mit der tatsächlichen Funktionsweise des Systems ein einklang gebracht.

Auf Basis der Ansätze, Erklärungen zu definieren, wurde *Explainability* als neue Nicht-Funktionale Anforderung eingeführt [19], um den Einsatz von Erklärungen und die Abhängigkeiten mit anderen NFRs besser untersuchen zu können.

2.1.1 Erklärbarkeit als Nicht-Funktionale Anforderung

Nicht-Funktionale Anforderungen sind Anforderungen an ein oder mehrere Qualitätsaspekte im Rahmen der Softwarequalität [31, 32]. Dabei ist Softwarequalität laut *IEEE 1061-1992 - IEEE Standard for a Software Quality Metrics Methodology* als „[...] the degree to which software possesses a desired combination of attributes“[33] definiert. Folglich müssen für Erklärbarkeit Attribute bestimmt werden, welche einen Einfluss auf die Erfüllung dieser NFR haben.

Eine der ersten Interpretationen von *Explainability* als NFR liefern Köhl et al. Dabei gilt ein System als erklärbar, wenn es ein Mittel gibt, welches eine Erklärung generiert, die die beschriebenen Bedingungen erfüllt [19]. Konkrete Anforderungen an Erklärbarkeit für ein System müssen laut Köhl et al. außerdem eine bestimmte Nutzergruppe, einen Kontext sowie den zu erklärenden Aspekt des Systems enthalten.

Auf Basis der bereits erfolgten Definitionen für Erklärbarkeit als Nicht-Funktionale Anforderung sowie einer systematischen Literaturrecherche haben Chazette, Brunotte und Speith die neue NFR im Kontext von *Requirements Engineering* folgendermaßen definiert:

> A system **S** is explainable with respect to an aspect **X** of **S** relative to an addressee **A** in context **C** if and only if there is an entity **E** (the explainer) who, by giving a corpus of information **I** (the explanation of **X**), enables **A** to understand **X** of **S** in **C** [5].

Der *Addressee (A)* ist dabei der Stakeholder, welcher eine Erklärung erhalten soll. Der *Context (C)* wird durch die Situation gegeben, welche durch die Interaktion eines Nutzers, seiner Aufgabe, dem System und der Umgebung entsteht. Was Köhl et al. in der oberen Definition als *Mittel* bezeichnet hat, beschreiben Chazette, Karras und Schneider als aktiven *Explainer*, welcher in Bezug auf ein System oder Systemteile Informationen an die *Addressees* gibt.

Mit dieser Definition stellen Chazette, Brunotte und Speith das Verständnis der *Addressees* in den Mittelpunkt von Erklärungen. Das heißt, sie betonen, dass die Nutzer eines erklärbaren Systems die Informationen nicht nur erhalten, sondern auch verstehen müssen, um die Lücke zwischen *Mental Model* und wirklicher Funktionsweise des Systems zu schließen.

Anschließend wird mit dieser Definition eine Verbindung zu bereits bestehenden Qualitätsaspekten hergestellt. [7]. Durch bereits bekannte Qualitätsaspekte – hier *Understandability* – kann Erklärbarkeit messbar gemacht werden. Durch die

Verknüpfung mit existierenden Qualitätsaspekten wird es möglich, auf bestehende Kataloge zur Messung dieser zurückzugreifen.

Neben *Understandability* wurden in vorangegangenen Arbeiten auch Abhängigkeiten zu weiteren Qualitätsaspekten wie z. B. *Transparency* untersucht [21]. Eine Übersicht, mit welchen Qualitätsaspekten *Explainability* Wechselwirkungen haben, wird von Chazette, Brunotte und Speith in ihrer Arbeit beschrieben [5]. Ein Resultat ist, dass *Explainability* mit sehr vielen Aspekten in Verbindung steht. Grundsätzlich können dabei Qualitätsaspekte, die im Zusammenhang mit *Explainability* stehen, als Qualitätsziele für das Integrieren von Erklärungen gesetzt werden [21].

Durch den komplexen Zusammenhang von Softwarequalität mit *Explainability* ergeben sich allerdings auch konkurrierende Ziele [7]. So können Erklärungen unter anderem einen positiven Einfluss auf die *Transparency* eines Systems haben und gleichzeitig die Performanz der Systemnutzer verschlechtern. Dieser Trade-off muss für die Verbesserung der Softwarequalität durch die Integration von Erklärungen in ein System beachtet werden [vgl. 21].

Folglich müssen die Einflüsse von *Explainability* auf andere Qualitätsaspekte genauer untersucht werden. Außerdem müssen für die Überprüfung des Erfolges der Integration von Erklärung die Anforderungen an diese zuvor klar dargestellt werden.

2.2 Qualitätsmodelle

Eine gängige Möglichkeit konkrete Anforderungen zu formulieren bieten Qualitätsmodelle [32]. Die Idee dieser Modelle ist von allgemein formulierten Zielen, iterativ überprüfbare, konkrete Anforderungen zu entwickeln. Abbildung 2.1 definiert die drei Ebenen von Qualitätsmodellen: *Abstrakt*, *Konkret* und *Messbar*. In der Regel werden dabei definierte Qualitätsaspekte aus Normen als allgemeine Ziele formuliert.

Daraufhin werden diese für einen bestimmten Kontext konkretisiert, um ein besseres Verständnis dafür zu erlangen, worauf sich ein Qualitätsziel in dem spezifischen Anwendungsfall bezieht.

Aus den konkreten Zielen können dann Metriken abgeleitet werden, anhand derer festgelegt werden kann, ob das Ziel erfüllt ist. Dazu wird eine Anforderung formuliert, welche im Regelfall Sollwerte für die festgelegten Metriken enthält.

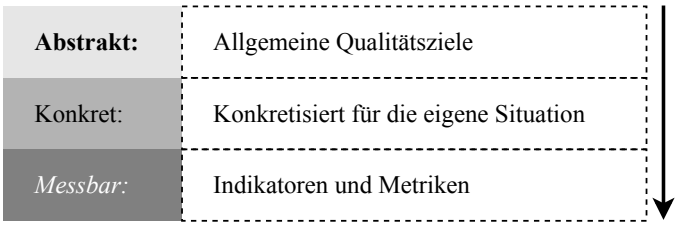


Abbildung 2.1: Aufbau eines Qualitätsmodells in drei Ebenen [S. 34, 32]

Diese Arbeit entwickelt zwar kein Qualitätsmodell für *Explainability*, enthält allerdings Zielvorschläge für die abstrakte Ebene des Modells. Im letzten Teil wird diese Basis verwendet, um Anforderungen für die Integration von Erklärungen in einem Beispielsystem aufzustellen. Welches Ziel diese Arbeit dabei verfolgt, wird im nächsten Kapitel vorgestellt.

Kapitel 3

Forschungsziel- und Design

In diesem Kapitel werden das Ziel und der Zweck dieser Arbeit beschrieben, die daraus abgeleiteten Forschungsfragen erläutert und in einen Zusammenhang gestellt. Des Weiteren wird das Vorgehen zur Beantwortung der Forschungsfragen beschrieben. Dabei ist das Forschungsdesign in eine Literaturrecherche, die Strukturierung der extrahierten Ergebnisse und die Anwendung des aus den Ergebnissen entwickelten Leitfadens gegliedert.

3.1 Zieldefinition

Die Zieldefinition dieser Arbeit erfolgt in Anlehnung an die Vorlage zur Formulierung von Zielen für Experimente von Wohlin et al. [34].

> Die Arbeit **analysiert** den Einsatz, die Gestaltung und die Evaluation von Erklärungen in Software-Systemen **in Bezug auf** die externe Qualität **zur** Entwicklung eines Leitfadens als Unterstützung für die Gestaltung von Erklärungen **aus der Sicht** von Software-Entwicklern **im Kontext** einer Literaturrecherche und anschließender Evaluation in der Praxis.

3.2 Forschungsfragen

Aus dem Forschungsziel wurden fünf Forschungsfragen (*Research Questions, RQ*) abgeleitet, welche die Richtung der Arbeit fein granularer definieren. Jede Forschungsfrage behandelt entweder einen konkreten Betrachtungsgegenstand von Erklärbarkeit oder den Zusammenhang der einzelnen Aspekte. Wie die einzelnen Forschungsfragen zusammenhängen, ist in Abbildung 3.1 dargestellt.

> **RQ1** Welche Rahmenbedingungen und Ziele haben einen Einfluss auf die Anforderungen für Erklärungen?

Erklärbare Systeme können sehr verschiedene Ausprägungen haben (z.B. Empfehlungssysteme [35] oder autonome Fahrzeuge [36]). Auch werden je nach Anwendungsgebiet verschiedene Anforderungen an die Systeme gestellt oder es gibt zusätzlich geltende, äußere Bedingungen [5].

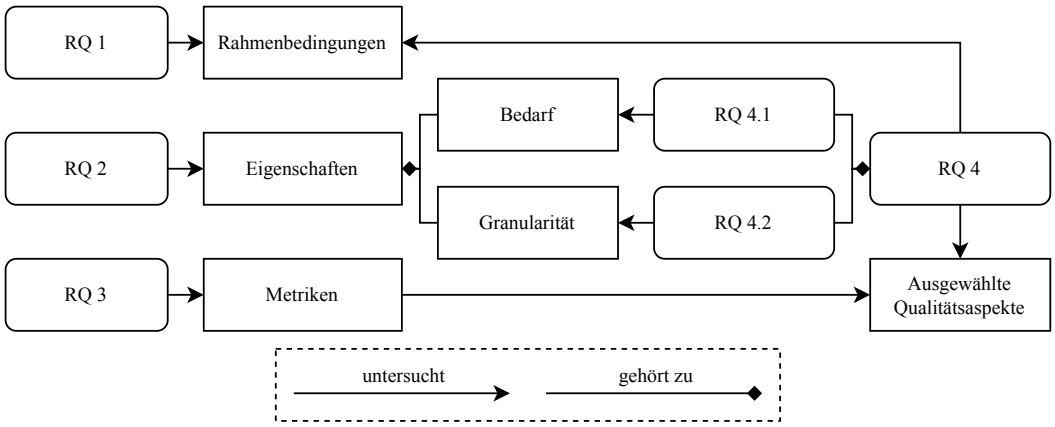


Abbildung 3.1: Zusammenhänge der Forschungsfragen

Die Untersuchung dieser Aspekte auf die Einflüsse auf die benötigten Erklärungen bzw. Erklärbarkeit im Allgemeinen geschieht als Antwort auf diese Frage. Dabei sollen Rahmenbedingungen ausgearbeitet werden, welche für Anforderungen an Erklärungen relevant sind. Auch mögliche Ziele für Erklärungen soll die Antwort auf diese Forschungsfrage bieten.

RQ2 Welche Eigenschaften von Erklärungen haben einen Einfluss auf die externe Qualität eines erklärbaren Systems?

Es existieren bereits verschiedene Frameworks, die für bestimmte Anwendungsbereiche die Gestaltungsmöglichkeiten für Erklärungen zusammenfassen [17]. Vor allem im Bereich der Künstlichen Intelligenz gibt es zahlreiche Arbeiten, welche einen solchen Überblick geben, um automatisch Erklärungen für komplexe Algorithmen zu generieren [37, 38].

Um der Forderung nach einem stärkeren Fokus auf den Menschen bei der Betrachtung von Erklärbarkeit nachzukommen [39], stellt diese Forschungsfrage die externen Qualitätsaspekte [6] in den Mittelpunkt. Hierbei werden externe Qualitätsaspekte mit einbezogen, durch welche die Qualität von Erklärungen bestimmt werden kann. Außerdem werden im Rahmen dieser Frage unter anderem die Gestaltungsmöglichkeiten für den Bedarf an Erklärungen und deren Granularität betrachtet, die Entwicklern oder Designern bei der Konzeption von Erklärungen zur Verfügung stehen.

RQ3 Auf welche Art und Weise kann evaluiert werden, ob die in ein erklärbares System integrierten Erklärungen das Ziel der Integration bezogen auf externe Qualitätsaspekte erfüllt haben?

Erklärbarkeit ist eine NFR, welche von vielen Faktoren abhängt und diese auch beeinflusst [5]. Die Integration von Erklärungen hat in der Regel Effekte auf mehrere andere Qualitätsaspekte. Diese Effekte können sowohl positiv als auch negativ sein. Durch die Integration von Erklärungen ist es unter anderem möglich, dass die *Usability* des Systems verschlechtert wird, wenn sie nicht explizit mitgedacht wird [37]. Folglich stellt das Messen der Qualität von Erklärungen aufgrund der vielfältigen Effekte eine Herausforderung dar. Des Weiteren muss bei der Evaluation beachtet werden, dass viele der durch Erklärbarkeit beeinflussten NFRs vorwiegend

subjektiv von Software-Nutzern wahrgenommen werden und somit nur wenige objektive Metriken einsetzbar sind [37].

In der Literatur wurden bereits Erklärungen anhand verschiedener Metriken analysiert [36, 40]. Es fehlt allerdings ein Überblick, welche Metriken zum Messen und zur Bewertung der Qualität von Erklärungen geeignet und erprobt sind. Folglich ist das Ziel bei der Beantwortung dieser offenen Frage, die bereits zur Evaluation von Erklärungen genutzten Metriken zusammenzutragen.

RQ4.1 Welchen Einfluss hat der Bedarf von Erklärungen auf die externe Qualität eines erklärbaren Systems unter bestimmten Rahmenbedingungen?

RQ4.2 Welchen Einfluss hat die Granularität von Erklärungen auf die externe Qualität eines erklärbaren Systems unter bestimmten Rahmenbedingungen?

In den ersten drei Forschungsfragen wurden verschiedene Aspekte von Erklärbarkeit behandelt, die wichtig sind, um darauf basierend Erklärungen in ein System zu integrieren. Allerdings werden durch die Antworten auf die Fragen keine Empfehlungen für die richtige Auswahl der Aspekte gegeben. Diese Aufgabe soll durch die Antwort auf diese letzten beiden Fragen unterstützt werden.

Die Einflüsse, nach denen RQ4.1 fragt, sollen bei der Einschätzung helfen, ob und wenn ja, an welchen Stellen, ein System Erklärungen benötigt. Dies soll Entwicklern und Designern dabei helfen, die Frage nach dem Bedarf für Erklärungen Anwendungsfall-übergreifend zu beantworten. Für Navigationsanwendungen haben dies beispielsweise Chazette, Karras und Schneider sowie Wang bereits getan [7, 21].

Wenn eben dieser Bedarf besteht, muss anschließend die Ausgestaltung dieser Erklärung betrachtet werden. Welche bereits untersuchten Einflüsse in Bezug auf diese Granularität von Erklärungen bestehen, soll die Antwort auf die Frage RQ4.2 beantworten.

Folglich stellen die Forschungsfragen RQ4.1 und RQ4.2 die Schnittstelle zwischen den ersten drei Forschungsfragen dar.

3.3 Forschungsdesign

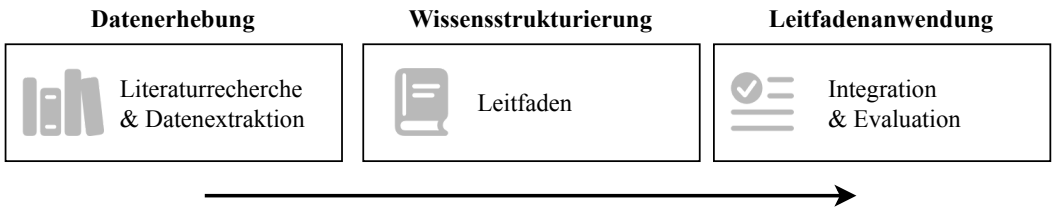


Abbildung 3.2: Überblick des Forschungsdesigns

Eine Übersicht der Vorgehensweise zur Beantwortung der Forschungsfragen ist in Abbildung 3.2 dargestellt. Der Forschungsansatz besteht aus zwei wesentlichen zusammenhängenden Teilen.

Zunächst wurden Daten zum aktuellen Wissensstand über die Aspekte von Erklärungen sowie dessen Zusammenhänge und Auswirkungen auf Softwarequalität

erhoben. Zu diesem Zweck wurde eine Literaturrecherche mittels der Suchstring-Methode [41] durchgeführt. Die Antworten auf die Forschungsfragen sind aus den selektierten Arbeiten extrahiert worden, wobei zunächst die Forschungsfragen RQ1 bis RQ3 bearbeitet wurden, um darauf aufbauend die Fragen nach den Einflüssen (RQ4.1 und RQ4.2) zu beantworten.

Auf Basis der Resultate dieser Datenerhebung konnten die Ergebnisse zu den ersten drei Forschungsfragen in ein Modell über die Aspekte von Erklärbarkeit überführt werden. Dies orientiert sich sowohl an Strukturen, welche die Literaturrecherche ergeben hat, als auch an den Forschungsfragen und der Definition von Erklärbarkeit von Chazette, Brunotte und Speith [5] (siehe Abschnitt 5.2).

Die erhobenen Daten zur Beantwortung der Forschungsfragen RQ4.1 und RQ4.2 sind in einem Katalog der Zusammenhänge zwischen Rahmenbedingen, Charakteristiken von Erklärungen und dem Einfluss auf ausgewählte für Endnutzer wahrnehmbare Qualitätsaspekte (externe Qualitätsaspekte) gebündelt. Als Schlussfolgerung daraus wurden zusätzlich Design-Empfehlungen abgeleitet.

Der zweite Teil der Forschung besteht aus der Anwendung des entstandenen Leitfadens. Dies dient der Prüfung, ob der Leitfaden die Integration von Erklärungen in ein bestehendes System unterstützt. Das Ziel ist es, durch neu integrierte Erklärungen positive Auswirkungen auf zuvor definierte Qualitätsziele zu erhalten.

Die Prüfung der Praxistauglichkeit des Leitfadens wurde zusammen mit dem Unternehmen *Graphmasters GmbH* [1] aus Hannover durchgeführt. *Graphmasters* ist unter anderem das entwickelnde Unternehmen hinter der Navigationssoftware *NUNAV Navigation* [2], welches ein Smartphone-Navigationssystem für Endnutzer mit einem kollaborativen Routing-Ansatz ist (Für mehr Details siehe Kapitel 6).

Zunächst wurden bestehende Nutzungsprobleme von *NUNAV* analysiert. Darauf basierend hat ein interdisziplinäres Team von *Graphmasters* in einem vorbereiteten Workshop (siehe Abschnitt B) anhand des zuvor entwickelten Leitfadens grundlegende Ideen für Erklärungen gesammelt. Mithilfe dieser konnten konkrete Anforderungen abgeleitet werden.

Auf Basis der Ergebnisse wurden dann mehrere Erklärungstypen entwickelt und in NUNAV integriert. Diese sind im nächsten Schritt in einer *Case Study* mit Nutzern des Systems evaluiert worden.

Als Abschluss der Forschung wurde zur besseren Interpretation der qualitativen Ergebnisse der *Case Study* ein nicht repräsentatives qualitatives Experiment (Quasi-Experiment) unter Nutzern von *NUNAV Navigation* durchgeführt.

In den folgenden Kapiteln werden die einzelnen beschriebenen Schritte zusammen mit den Ergebnissen näher erläutert.

[1] https://www.graphmasters.net, besucht: 10.09.21
[2] https://www.nunav.net, besucht: 10.09.21

Kapitel 4

Literaturrecherche

Die Grundlage für das weitere Vorgehen bildet eine Literaturrecherche. Das Ziel dabei ist es, einen Überblick über bereits evaluierte Erklärungstypen in verschiedenen Kontexten zu erhalten. Aus den Typen sowie unterschiedlichen Evaluationsarten kann dann der zu entstehende Leitfaden abgeleitet werden.

Die Literaturrecherche besteht aus einer Planungs- und einer Durchführungsphase. In der Planungsphase wurden die Rahmenbedingungen dafür festgelegt, welche vorangegangen Arbeiten für das Ableiten eines Leitfadens zur Integration von Erklärungen geeignet sind. Anschließend werden in der Durchführungsphase alle daraus resultierenden Veröffentlichungen weiter gefiltert und für den Leitfaden zur Integration von Erklärungen wichtigen Informationen extrahiert.

4.1 Planung

Als Methode für die Literaturrecherche ist die Suchstring-Methode zur Anwendung gekommen. Dabei werden am Anfang die Datenbanken und Suchbegriffe definiert, um eine initiale Menge an Suchergebnissen zu erhalten.

Als Datenbanken wurden für die Suche *ACM Digital Libraray*[1], *IEEE Xplore*[2], *Science Direct*[3] sowie *Springer Link*[4] verwendet. Die genannten Datenbanken wurden gewählt, da sie bereits für Literaturrecherchen im Bereich von Erklärbarkeit eingesetzt wurden [17, 42]. Bei der Suche in *Science Direct* und *Springer Link* wurden die Ergebnisse auf den Bereich *Computer Science* beschränkt. Weitere Filtereinstellungen für jede Datenbank sind in Tabelle A.1 zu finden.

Als Zeitraum für die Veröffentlichungen wurde das Jahr 2015 bis zur Durchführung der Suche (08.06.21) gewählt. 2015 wurde dabei als Startjahr gewählt, da zu diesem Zeitpunkt die Zahl der Veröffentlichungen zum Thema Erklärbarkeit mit Fokus auf die Wahrnehmung von Nutzern deutlich ansteigt. Eine zusätzliche Betrachtung der ersten Phase von Veröffentlichungen zu Erklärbarkeit mit Blick auf Psychologie um 1990 würde den Rahmen dieser Arbeit überschreiten.

[1] https://dl.acm.org
[2] https://ieeexplore.ieee.org
[3] https://www.sciencedirect.com
[4] https://link.springer.com

4.1.1 Definition der Suchbegriffe

Bei der Definition der Suchbegriffe wurden zunächst Schlüsselbegriffe definiert. Dabei wurden drei Blöcke von Begriffen identifiziert. Zur Themenabgrenzung muss jedes Suchergebnis ein Synonym für „Erklärung“ oder „Erklärbarkeit“ aufweisen. Da die vorliegende Arbeit unter anderem Einflüsse und Evaluationen untersucht, besteht der zweite Begriffsblock aus solchen der Wortfamilie „Evaluation“. Aus der Betrachtung von externer Qualität wurden als letzter Block Begriffe, welche mit dem Gebiet der Mensch-Maschine Kommunikation in Verbindung stehen, gewählt. Eine vollständige Übersicht der Synonyme ist in Tabelle 4.1 zu sehen.

Erklärbarkeit	Evaluation	Mensch-Maschine Kommunikation
explainability	evaluation	HCI
explanation	assessment	human-computer interaction
explanations	analysis	human-computer interfaces
explainable	impact	interaction
		user interface
		usability

Tabelle 4.1: Schlüsselbegriffe für die Konstruktion des Suchstrings

Der finale Suchstring ist so aufgebaut, dass zusätzlich zur ersten Bedingung entweder ein Begriff aus dem zweiten oder dritten Block von Begriffen auf ein Suchergebnis zutreffen muss:

((explainability OR explanation OR explanations OR explainable) AND (evaluation OR assessment OR analysis OR impact OR HCI OR „human-computer interaction“ OR „human-computer interfaces “ OR interaction OR „user interface“ OR usability))

Zur Kontrolle, ob die Suche relevante Ergebnisse liefert wurden vier bekannte Arbeiten verwendet, von denen bekannt war, dass diese Antworten auf die Forschungsfragen enthalten: Chazette, Karras und Schneider [7], Chazette und Schneider [2], Köhl et al. [19], Sokol und Flach [37].

4.1.2 Auswahlkriterien für Primärliteratur

Um Suchergebnisse auf die für das Ziel der dieser Arbeit zu verwendenden Ergebnisse zu begrenzen wurden Bedingungen für das Selektieren (*Inclusion Criteria: IC*) und Ausschließen (*Exculsion Criteria: EC*) gewählt:

Bedingungen

IC1 Die Arbeit muss *peer reviewed* sein.

IC2 Die Arbeit muss in Englisch oder Deutsch verfasst sein.

IC3 Die Arbeit muss entweder die Evaluation einer bestimmten Erklärung oder einen Überblick über verschiedene Evaluationsmöglichkeiten enthalten.

IC4 Die Arbeit muss End-Nutzer von erklärbaren Systemen als Stakeholder von Erklärungen in Betracht ziehen.

Ausschlüsse

EC1 Die Arbeit darf sich nicht ausschließlich darauf fokussieren, wie Erklärungen automatisch generiert werden können (Algorithmus-Evaluation).

EC2 Die Arbeit darf sich nicht ausschließlich auf das Verstehen von zugrundeliegenden Algorithmen beschränken (ML-*Interpretability*).

4.2 Durchführung

Nach der initialen Suche wurden mehrere Filterschritte durchgeführt. Abbildung 4.1 bietet einen Überblick über das Auswahlverfahren der Arbeiten, auf denen die hier entstandenen Ergebnisse aufbauen. Der Auswahlprozess ist in drei Iterationen erfolgt. Die Suche in den vier vorgestellten Datenbanken hat insgesamt 790 Ergebnisse ergeben. Dabei gab es 141 Duplikate. Die initiale Menge nach der Suche enthielt folglich 649 wissenschaftliche Arbeiten. Da die Suche auf die genannten vier Datenbanken beschränkt und passende Filter genutzt wurden, erfüllen diese Veröffentlichungen bereits die Kriterien des *Peer Reviews*([IC1]).

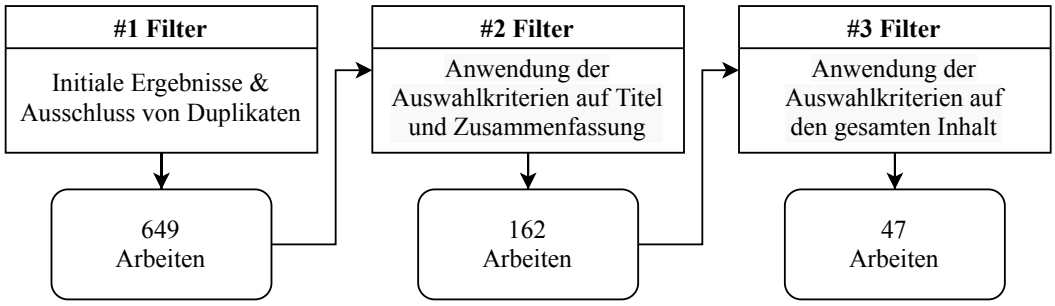


Abbildung 4.1: Auswahlverfahren der Literaturrecherche

In einem zweiten Schritt wurden die Selektionskriterien auf den Titel und die Zusammenfassung der Arbeiten angewendet. Für die übrigen 162 Veröffentlichungen wurde der Volltext auf die Kriterien überprüft, wobei am Ende 47 Arbeiten alle Selektionskriterien erfüllt haben. Im Anschluss an die Auswahl wurden neun weitere Ergebnisse aufgenommen, die zum Verständnis einer anderen Arbeit wichtig waren und ebenfalls die Selektionskriterien erfüllt haben. Final enthält die Menge der wissenschaftlichen Arbeiten, die als Informationsgrundlage für die Erstellung eines Leitfadens zur Integration von Erklärungen gedient hat, 56 Arbeiten.

4.3 Ergebnisse

Die Ergebnisse der Literaturrecherche lassen sich zunächst in verschiedene Kontexte einordnen (siehe Abbildung A.1). Dabei werden die Bereiche Intelligente Systeme (z.B. XAI), Empfehlungssysteme, Autonomes Fahren sowie Mensch-Roboter Interaktion abgedeckt. Außerdem betrachten einige Arbeiten Erklärbarkeit im Allgemeinen. Ferner gibt es zwei Arbeiten die jeweils eine sehr Domänen- oder Anwendungsspezifische Evaluation durchführen.

Ein wichtiges Selektionskriterium war, dass die Arbeiten entweder eine Evaluation von Erklärungen enthalten oder Möglichkeiten der Evaluation zusammenfassen. Daher ist als zweites wichtiges Merkmal der Ergebnisse die empirische Strategie zu betrachten. Tabelle 4.2 enthält eine Übersicht über die verschiedenen Ziele der Arbeiten sowie verwendeten empirischen Strategien.

Bei Evaluationen werden in der Literatur vorwiegend kontrollierte Experimente, sowie vereinzelt Case Studies und Umfragen verwendet. Bei der Analyse bestehender Ergebnisse werden in den resultierenden Arbeiten entweder Literaturrecherchen oder Umfragen eingesetzt.

Auf Basis der finalen Suchergebnisse konnten aus den Arbeiten die Antworten auf die Forschungsfragen als zunächst als Rohantworten extrahiert werden. Diese Datengrundlage dient für den Aufbau des Leitfadens für die Integration von Erklärungen, welcher im folgenden Kapitel vorgestellt wird.

Ziel der Arbeit	Empirische Strategie	Quellen
Evaluation des Einflusses von Erklärungen auf bestimmte Qualitätsaspekte	Kontrolliertes Experiment	[43] [44] [45] [46] [47] [22] [35] [48] [49] [50] [51] [52] [29] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63]
	Umfrage	[7] [2] [64]
	Case Study	[65] [66]
Evaluation von Nutzer-Präferenzen für verschiedene Erklärungen	Kontrolliertes Experiment	[18] [67] [68] [69] [70] [71] [72] [36] [73]
Überblick über bestehende Ergebnisse	Literaturrecherche	[5] [37] [3] [19] [30] [74] [75] [76] [26] [77] [17] [23] [78] [79] [25] [80] [81]
	Umfrage	[20]

Tabelle 4.2: Ergebnisse der Literaturrecherche nach Art der Publikation

Kapitel 5

Leitfaden zur Integration von Erklärungen

Dieses Kapitel beschreibt den Leitfaden zur Integration von Erklärungen, welcher auf Basis der vorangegangenen Literaturrecherche entwickelt wurde (siehe Kapitel 4). Zuerst werden die konkreten Anforderungen vorgestellt, welche in vorangegangener Literatur erarbeitet wurden, die für einen Leitfaden dieser Art gelten. Neben einem Überblick über die verschiedenen Einflussfaktoren, die bei der Entwicklung von Erklärungen für ein erklärbares System betrachtet werden sollten, gibt Abschnitt 5.2 außerdem einen Überblick über die Ausprägungen der Eigenschaften von Erklärungen, die in der Literatur bereits Anwendung gefunden haben. Die Evaluation von Erklärungen wird in Unterabschnitt 5.2.4 thematisiert. Abschnitt 5.3 enthält eine Übersicht von bereits gezeigten Zusammenhängen zwischen bestimmten Aspekten von Erklärungen basierend auf dem Modell der verschiedenen Aspekte von Erklärbarkeit. Abschließend werden in diesem Kapitel die Auswirkungen daraus für das Design von Erklärungen vorgestellt (Abschnitt 5.4).

5.1 Anforderungen an den Leitfaden

Für die Integration von Komponenten in ein System werden entsprechende Anforderungen dafür benötigt. Dazu verweisen Chazette, Karras und Schneider darauf, dass es einer Unterstützung bedarf, Anforderungen an Erklärungen zu formulieren. Hierzu gehört unter anderem die Identifikation von Problemen, welche durch Erklärungen gelöst werden können [7, 25]. Unterstützt wird dies durch Waa et al., welche dies nutzen wollen, um jeder Evaluation von Erklärungen klare Hypothesen voraus zustellen und somit die Ergebnisse verschiedener Arbeiten besser zusammenfassen und daraus Empfehlungen ableiten zu können [69]. Auch Köhl et al. unterstreichen den Aspekt, dass vor allem die Ziele und Anforderungen auf Basis des aktuellen Kontextes definiert sein müssen, bevor Erklärungen in ein System integriert werden können [19]. Die erarbeiteten Anforderungen an einen Leitfaden bilden die Grundlagen, um die gestellte Forschungsfrage nach dem Einfluss der Rahmenbedingungen eines erklärbaren Systems zu beantworten (RQ1, siehe Abschnitt 3.1).

Außerdem fordern Waa et al., dass zur Orientierung bei der Integration von Erklärungen ein Überblick über bereits entwickelte Erklärungstypen erstellt werden sollte. Zusammen mit RQ2 folgt daraus die Forderung nach einer Zusammenfassung der bestehenden Ergebnisse zum Einsatz verschiedener Erklärungstypen.

RQ3 fragt nach den Metriken, durch welche die Qualität von Erklärungen gemessen werden kann (Abschnitt 3.1). Nach unabhängigen Untersuchungen verschiedener Erklärungstypen stellen mehrere Autoren den Bedarf für eine Vereinheitlichung der Evaluation von Erklärungen fest [75, 29, 17, 55]. Dies soll es unter anderem ermöglichen, die Vergleichbarkeit unterschiedlicher Lösungsansätzen zu gewährleisten. In einigen Arbeiten fordern die Autoren dabei explizit, dass ein Framework zur Evaluation benötigt wird, anhand dessen die Qualität von Erklärungen bestimmt werden kann [17, 37, 77].

Außerdem ergibt sich sowohl aus dem Ziel der vorliegenden Arbeit (Abschnitt 3.1) sowie der Übersicht über bestehende Ergebnisse von Waa et al., dass der Leitfaden auch Gestaltungsempfehlungen für Erklärungen enthalten soll. Auch wird dies von Waa et al. im Rahmen einer Übersicht bestehender Ergebnisse gefordert [69].

Basis hierfür sind die Forschungsfragen RQ4.1 und RQ4.2, in welchen nach den Einflüssen auf verschiedene Eigenschaften von Erklärungen gefragt wird. Diese Einflüsse sollten so dargestellt werden, dass mögliche konkurrierende Qualitätsaspekte bei der Wahl verschiedener Eigenschaften für Erklärungen herausgestellt werden.

Zusammenfassend muss der Leitfaden mit dem Modell folgende Aspekte enthalten, um die Anforderungen aus der Literatur zu erfüllen und die Forschungsfragen zu beantworten (*Guideline Requirements (GR)*):

GR1 Der Leitfaden muss eine Unterstützung für das Erheben von Anforderungen an Erklärungen und das Aufstellen von Hypothesen über den Einfluss enthalten.

GR2 Der Leitfaden muss mögliche Vorschläge zur Umsetzung von Erklärungen geben.

GR3 Der Leitfaden muss einen Überblick über verschiedene Evaluationsmöglichkeiten für Erklärungen im Kontext externer Softwarequalität geben.

GR4 Der Leitfaden muss in der Literatur gezeigte Einflüsse einzelner Eigenschaften auf die externe Qualität von Erklärungen zusammenfassen.

Für die Umsetzung dieser Anforderungen ist der Leitfaden in zwei Teile gegliedert. Ersterer enthält ein Überblick über die verschiedenen Aspekte, die bei der Entwicklung von Erklärungen berücksichtigt werden sollten. Dieser Überblick wird in Form eines Modells gegeben, in dem die verschiedenen Aspekte gegliedert sind. Mithilfe des Modells soll die Möglichkeit geschaffen werden, Anforderungen und Hypothesen aufzustellen, diese umzusetzen und schlussendlich zu evaluieren.

Unterstützt werden soll die Entwicklung von Erklärungen außerdem durch bereits gezeigte Abhängigkeiten zwischen den im ersten Teil des Modells enthaltenen Aspekten sowie daraus abgeleiteten Auswirkungen auf das Design. Daher sind die beschriebenen Auswirkungen der zweite Teil des Leitfadens.

5.2 Aspekte von Erklärungen

Literatur, die einen Überblick über Erklärbarkeit im Allgemeinen oder in einem bestimmten Anwendungsfeld gibt, betrachtet meist fünf Aspekte von Erklärbarkeit [30, 17, 5]. Die fünf Aspekte setzen sich aus dem Kontext der Erklärung, der Zielsetzung dieser, der Granularität von Erklärungen und wann diese angezeigt werden sollen, zusammen. In einigen Arbeiten wird bei der Granularität außerdem zwischen dem Inhalt und der Darstellung unterschieden oder diese Eigenschaften einzeln untersucht [17, 68]. Zudem wird in allen hier betrachteten Arbeiten auch die Evaluation von Erklärungen thematisiert (siehe Kapitel 4).

Für die zuvor genannten Aspekte werden in der Literatur verschiedene Unterkategorien konkret benannt oder Synonyme verwendet. Tabelle 5.1 fasst die verwendeten Synonyme aus den Veröffentlichungen, welche den Aspekt explizit erwähnt haben, unter der final gewählten Benennung zusammen. Neben den dort aufgezeigten Begriffen haben mehrere Autoren (z. B. [30, 2]) die verschiedenen Aspekte von Erklärbarkeit zusätzlich mit den zuvor herausgearbeiteten Fragewörtern verknüpft. Die vollständigen Fragen dahinter verweisen allerdings auf verschiedene Unterpunkte, weswegen das vorgestellte Modell auf Fragewörter verzichtet, um Verwechselungen vorzubeugen. (Beispiel: „**Wie** kann die Erklärung evaluiert werden?“ (*Evaluation*) [vgl. 30] und „**Wie** viele Informationen sollte jede Erklärung enthalten?“ (*Content*) [vgl. 18]).

Im Folgenden werden die genannten Oberkategorien erläutert. Da alle im Leitfaden definierten Begriffe bereits in Englisch vorlagen, sind diese zur Erhöhung der Wiederverwendbarkeit in der Originalsprache verblieben.

Context Der *Context* einer Erklärung wird durch die Situation gegeben, welche durch die Interaktion eines Nutzers, seiner Aufgabe, dem System und der Umgebung entsteht. ([vgl. 5, 19]).

Objectives *Objectives* sind die Qualitätsziele, welche für eine Erklärung gelten oder aufgrund derer Erklärungen in ein System integriert werden sollen.

Demand Der *Demand* für eine Erklärung ist der Bedarf für Erklärungen durch den Nutzer eines Systems. Das heißt der *Demand* beschreibt die Notwendigkeit, zu einem bestimmten Zeitpunkt für einen Systemteil Erklärungen bereitzustellen. Das beinhaltet auch, ob die Initiative vom Nutzer ausgeht oder das System selbstständig eine Erklärung anzeigt.

Content Der *Content* einer bei bestehendem Bedarf dem Nutzer bereitgestellten Erklärung ist durch die Informationen und die Informationsdichte definiert.

Presentation Der Inhalt einer Erklärung kann Nutzern auf verschiedenen Wegen zugänglich gemacht werden. Die *Presentation* einer Erklärung ist die Art und Weise, auf die dem Nutzer die Informationen durch die Erklärung bereitgestellt werden.

Aspekt	Synonyme	Quellen
1. Context	(Experimental) Context	[5] [7] [44] [69] [19] [57] [23] [25]
	(Explanation) Scope	[34] [45] [25]
	Use Case	[69]
	Stakeholder	[30]
2. Objectives	Objectives	[17]
	Construct	[69]
	Purpose	[17] [34]
	(Stakeholder) Goals	[75] [23] [78]
	Main Drive	[82]
	Intended Effect	[22]
3. Demand	Demand	[5]
4. Content	User Interface Component(s)	[17] [30]
	Content	[78]
	Granularity	[5] [19]
5. Presentation	Presentation	[30] [18]
	(Explanation) Type	[78] [30]
6. Evaluation	Evaluation	[19] [25]
	Measurements	[69] [22]
	Metrics	[17] [82] [77] [69]

Tabelle 5.1: Übergeordnete Aspekte von Erklärungen, die in einschlägiger Literatur untersucht wurden

Evaluation Die *Evaluation* ist die Bewertung der Qualität von Erklärungen. Dies enthält die grundsätzlichen verschiedenen Evaluationsmöglichkeiten und -metriken, mit denen die Qualität gemessen werden kann.

5.2.1 Struktur des Modells

Die Struktur des Modells für Erklärungen orientiert sich an den Forschungsfragen und Anforderungen an das Modell, Ergebnissen aus der Literaturrecherche sowie dem bestehenden Konzept der Qualitätsmodelle [32] (siehe Abschnitt 2.2).

In der Literatur wurden einige Aspekte des Modells nicht nur verschieden benannt, sondern auch zum Teil in Oberkategorien gegliedert oder auf verschiedenen Abstraktionsebenen betrachtet. Hieraus ergibt sich eine hierarchische Anordnung des Modells. Verschiedene Abstraktionsebenen enthalten auch die von Schneider vorgestellten Qualitätsmodelle [32]. Dabei werden abstrakte Ziele (*Objectives*) immer weiter konkretisiert, bis sie schlussendlich mit konkreten Metriken messbar sind. Außerdem wird in dem zugrundeliegenden Modell („Goal-Driven and Property-Based Definition Approach for Product Metrics“ [40]) von Briand, Morasca und Basili definiert, dass die nötigen Abhängigkeiten der *Objectives* von äußeren

Faktoren, den im hier vorgestellten Modell für Erklärungen erwähnten *Context* entsprechen.

Die Betrachtung von *Context* und *Objectives* erfolgt in der Literatur über Erklärungen zum Teil in verschiedener Reihenfolge. Rosenfeld und Richardson schreiben, dass die erste Frage, welche geklärt werden sollte, die Frage „Warum benötigt das System eine Erklärung?“ ist [vgl. S. 699 30][17]. Im Gegensatz dazu schreiben Cirqueira et al., dass zuerst äußere Umstände, wie der Endnutzer des Systems geklärt sein sollten („Stakeholder Setting“ [75]), um darauf aufbauend die Ziele festzulegen. Hieraus resultiert, dass die Festlegung der Ziele mit ihren Abhängigkeiten wie auch von [32] geschildert eine iterative und stark zusammenhängende Aufgabe ist. Daher werden die Punkte *Context* und *Objectives* aus Tabelle 5.1 unter *External Dependencies* zusammengefasst.

Analog werden *Demand*, *Content* und *Presentation* unter einem gemeinsamen Aspekt gegliedert (*Characteristics*), da sich alle drei Aspekte direkt auf die Eigenschaften von Erklärungen beziehen. Damit wird der Zusammenhang zwischen den verschiedenen Merkmalen einer Erklärung verdeutlicht [17].

Neben der Evaluation der Qualitätsziele für Erklärungen betrachtet das hier vorgestellte Modell außerdem die unmittelbare Evaluation der Eigenschaften von Erklärungen.

Schlussendlich ist das Modell in die drei Oberkategorien *External Dependencies*, *Characteristics* und *Evaluation* gegliedert. Eine Übersicht des Modells für Erklärungen ist in Abbildung 5.1 zu sehen. Diese enthält darüber hinaus die in den folgenden Abschnitten beschriebenen Ausprägungen der einzelnen Kategorien aus der Literatur. Die Gesamtübersicht ist grafisch an der Taxonomie für Erklärungen von Nunes und Jannach angelehnt [17]. Die erwähnte Taxonomie ist allerdings nur auf den Einsatz von Erklärungen in Empfehlungssystemen bezogen und beschränkt sich daher u. a. auf bestimmte Darstellungstypen. Sie kann folglich nicht ohne Weiteres in andere Kontexte übertragen werden. Ebenso fehlt der Aspekt der Evaluation. Dieser wird allerdings nicht nur von Nunes und Jannach selbst, sondern auch von weiteren Autoren für wichtig erachtet [75, 55, 17].

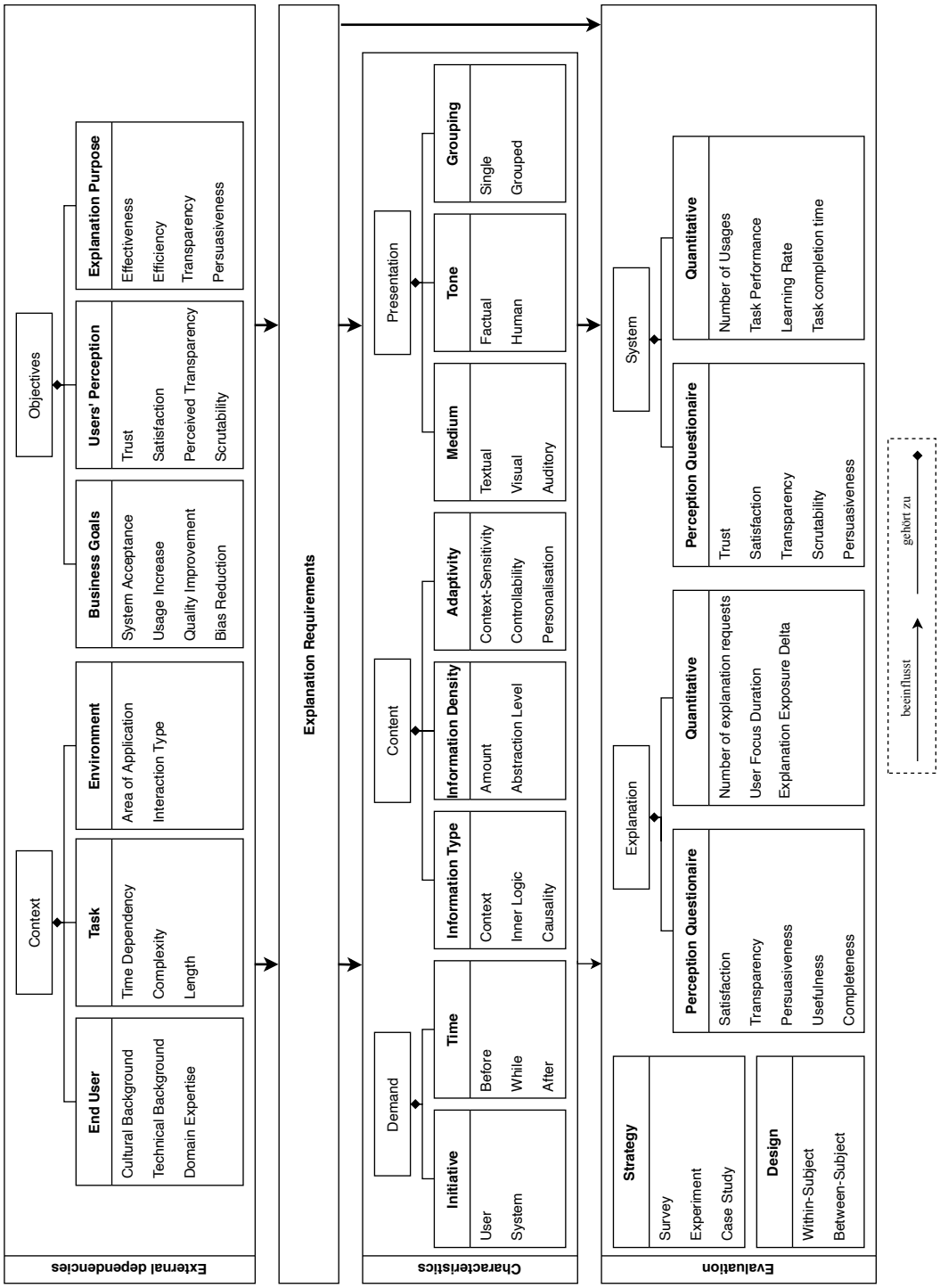


Abbildung 5.1: Übersicht über die Aspekte von Erklärbarkeit sowie bekannte Ausprägungen aus der Literatur

5.2.2 Externe Abhängigkeiten

Unter *External Dependencies* sind die Aspekte zusammengefasst, die eine Auswirkung auf die Erklärungen in einem System haben. Aus den hier vorgestellten Aspekten können im Anschluss Anforderungen und Einflusshypothesen für Erklärungen aufgestellt werden. Daraus kann dann auch abgeleitet werden, welche Funktionen des Systems einer Erklärung bedürfen [19]. Im Folgenden werden die beiden zusammenhängenden Unteraspekte *Context* des Systems und *Objectives* erläutert.

Context

Der *Context* einer Erklärung beschreibt die äußeren Einflüsse, die unmittelbar auf das erklärbare System wirken und aus denen somit Anforderungen an die Eigenschaften von Erklärungen abgeleitet werden können. Einen Überblick über den *Context* bietet Abbildung 5.2.

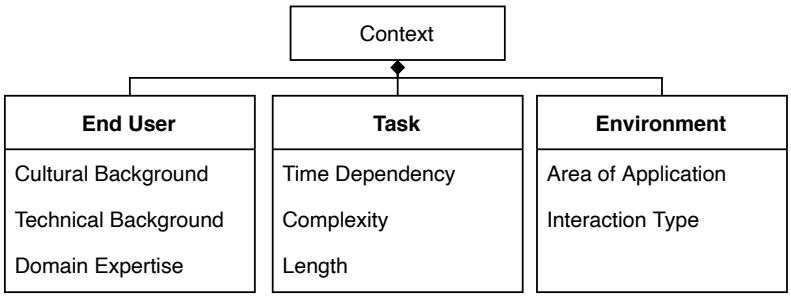


Abbildung 5.2: Übersicht über den *Context* eines erklärbaren Systems

Dies beinhaltet die Aktivität, die der Endbenutzer in einer bestimmten Umgebung durchführt. Aus den Eigenschaften der drei Aspekte (Aktivität, Endbenutzer und Umgebung) leiten sich dabei direkte Einflüsse auf den Bedarf, den Inhalt und die Darstellung einer Erklärung ab. Tabelle 5.2 stellt die verwendeten Synonyme für die verschiedenen Facetten des *Context* eines Systems dar. Folgend werden nun die drei Aspekte sowie typische Ausprägungen oder Charakteristiken näher erläutert.

End User Der *End User* ist diejenige Person, die mit dem System interagiert und auf welchen somit die Erklärungen zugeschnitten sein müssen. Dieser entspricht in den Definitionen von Erklärbarkeit von Chazette, Brunotte und Speith sowie von Köhl et al. dem *Explainee* [5, 19]. Allerdings muss der *Explainee* nicht zwangsweise *End User* des Systems sein, sondern einer anderen Stakeholdergruppe angehören.

Generell kann ein System verschiedene Nutzer(-typen) haben, die sich in ihrem Bedarf für Erklärungen unterscheiden. Im Folgenden werden die am häufigsten erwähnten Eigenschaften vorgestellt (unter anderem in [5, 43, 50]).

Sowohl beim generellen technischen Verständnis (*Technical Background*) als auch beim Domänen-Wissen (*Domain Expertise*) [50] können *End User* verschieden viel Hintergrundwissen vorweisen. Somit können sich in einem System mit unterschiedlichen Eigenschaften der Nutzer auch verschiedene Anforderungen an Erklärungen ergeben.

Aspekt	Synonyme	Quellen
End User	(Targt / End) User	[2] [72] [64] [70]
	Stakeholder	[5]
	Consumer	[66]
	Explainee	[5] [19]
	Explanation Audience	[37]
Task	Task	[5] [37] [79]
	Activity	[34]
Environment	Environment	[5] [70] [36]
	Application Area	[37] [36] [70]

Tabelle 5.2: Relevante Aspekte des *Context* eines erklärbaren Systems zur Integration von Erklärungen

Der *Cultural Background* fasst darüber hinaus die kulturellen Hintergründe von *End Usern* zusammen, die sich zum Beispiel auf die Verwendung von Metaphern in Software bzw. Erklärungen auswirken können [83].

Task Der *Task* definiert die Aufgabe(n), welche durch die *End User* mithilfe des Systems durchgeführt werden sollen. Auch hier spielen verschiedene Eigenschaften eine Rolle. Genannt werden in der Literatur unter anderem die Zeitabhängigkeit (*Time Dependency*), die Komplexität (*Complexity*) und die Dauer der Aufgabe (*Length*). Die drei genannten Ausprägungen werden als wichtige Betrachtungsgegenstände für die Eigenschaften von Erklärungen beschrieben. Daraus resultierend sind auch diese Aspekte für die Anforderungserhebung im Kontext der Erklärbarkeit relevant [37].

Environment Sehr eng mit der zu erledigenden Aufgabe hängt auch dessen Umgebung zusammen. Das *Environment* ist durch die äußeren Umstände des Systems definiert. Dies beinhaltet den generellen Anwendungsbereich des Systems (*Area of Application*), welcher unter anderem die Kritikalität eines Systems definiert. Aber auch die Art der Interaktion zwischen *End User* und System fällt in diesen Bereich (*Interaction Type*) und hat eine Auswirkung auf die Anforderungen an Erklärungen [70].

Als Zwischenergebnis für **RQ1** kann festgehalten werden, dass der *Context* eines Systems eine der zu betrachtenden Rahmenbedingungen mit einem Einfluss auf die Anforderungen an Erklärungen ist.

Zielsetzung

Als zweiter Aspekt neben dem *Context* werden in der Literatur die *Objectives* für die Integration von Erklärungen mit einem Einfluss auf die Anforderungen an Erklärungen genannt [30, 17].

Chazette, Brunotte und Speith haben in ihrem Modell für Erklärbarkeit die Qualitätsaspekte zusammengefasst, die mit Erklärbarkeit in einem Zusammenhang stehen und daher als *Objectives* für die Integration von Erklärungen gesehen werden können. Zusammen mit weiteren Ergebnissen der Literaturrecherche haben sich die acht in Tabelle 5.3 aufgelisteten, externen Qualitätsaspekte herausgestellt, welche in der Regel auf die Auswirkungen durch Erklärungen untersucht wurden [17, 81]. Die Messergebnisse der Untersuchungen der Qualitätsaspekte wurden in den meisten Fällen genutzt, um darüber indirekt die Qualität der integrierten Erklärungen zu definieren. Tabelle 5.3 enthält eine Liste dieser externen Qualitätsaspekte zusammen mit deren Definitionen. Diese enthält alle Arbeiten, welche den jeweiligen Aspekt explizit untersuchen oder bestehende Ergebnisse dazu zusammenfassen.

In ihrer Definition von Erklärbarkeit sehen Chazette, Brunotte und Speith vor allem *Transparency* und *Understandability* als zentrale Ziele von Erklärbarkeit [7]. Sie schreiben, dass diese unmittelbar durch die Integration von Erklärungen erreicht werden können. *Understandability* ist dabei als „Das Verständnis von Nutzern über

Qualitätsaspekt	Beschreibung	Quellen
Trust	Das Vertrauen des Nutzers in das System erhöhen [vgl. 22]	[17] [5] [43] [22] [45] [3] [47] [71] [49] [50] [52] [65] [55] [56] [64] [59] [62] [36] [79] [80] [81] [35]
Satisfaction	Die Benutzerfreundlichkeit und generelle Zufriedenheit von Nutzern mit dem System erhöhen. [vgl. 22]	[17] [5] [43] [22] [46] [3] [54] [65] [55] [56] [66] [23] [62] [78] [79] [80] [81] [44]
Transparency	Erklären, wie das System funktioniert. [vgl. 22]	[17] [5] [43] [7] [22] [2] [3] [47] [56] [76] [64] [59] [62] [81] [55] [66] [63]
Scrutability	Nutzern die Möglichkeit geben, dem System einen Fehler mitzuteilen [vgl. 22]	[17] [5] [43] [22] [3] [65] [79] [81] [55]
Effectiveness	Die Qualität der Aufgaben von Nutzern erhöhen [vgl. 22]	[17] [5] [43] [22] [3] [53] [47] [55] [76] [81]
Efficiency	Nutzern helfen ihre Aufagaben schneller zu erledigen [vgl. 22]	[17] [5] [43] [22] [46] [3] [47] [81]
Persuasiveness	Die Akzeptanz der Entscheidungen des Systems durch die Nutzer erhöhen [vgl. 22]	[17] [43] [22] [44] [68] [3] [51] [81]

Tabelle 5.3: Untersuchte Qualitätsaspekte in der Literatur für die Qualitätsbestimmung von Erklärungen mit Definition

das System erhöhen“ definiert [vgl. 7]. Häufig werden die beiden Aspekte in der Literatur allerdings als Synonyme verwendet [17, 42, 43] oder zusammen evaluiert. Ferner ist *Understandability* als *Softgoal* für *Transparency* aufgeführt [24]. Daher fasst dieses Modell beide Aspekte unter *Transparency* zusammen. Allerdings wird im Rahmen von Erklärbarkeit wie in Unterabschnitt 2.1.1 beschrieben zusätzlich zwischen *Transparency* und *Perceived Transparency* unterschieden [17]. Letzteres ist dabei die wahrgenommene *Transparency* des *End Users*. Dies ist wichtig, da es möglich ist, dass nur die *Perceived Transparency* das Ziel der Integration von Erklärungen ist. Notwendig ist die Unterscheidung, da Unternehmen ggf. nicht bereit sind, die Funktionalität ihrer Software offenzulegen oder ein System, dessen Erklärungen eine hohe *Transparency* bieten, durch ihre Komplexität *End User* überfordern können [5].

Wie in Tabelle 5.3 zu sehen ist, werden die Qualitätsziele verschieden häufig untersucht. Dies ist ein Indiz dafür, dass die Aspekte einer Struktur bedürfen. Daher schlagen einige Autoren eine hierarchische Anordnung der Qualitätsziele vor [17, 81, 69]. Folglich werden die *Objectives* in drei Abstraktionsebenen gegliedert: *Business Goals*, *Users' Perception* und *Explanation Purpose*. Diese Ebenen, die von Nunes und Jannach sowie Tintarev und Masthoff vorgeschlagen werden, sind in diesem Modell als konkrete Abstraktionslevel für Qualitätsmodelle ([vgl. 32]) zu interpretieren. Die Zuordnung der Qualitätsaspekte zu den Kategorien ist in der Abbildung 5.3 abgebildet. In Tabelle 5.4 werden die Erwähnungen der drei Zielebenen mit den jeweiligen Synonymen aus der Literatur zusammengefasst.

Business Goals *Business Goals* sind die Ziele, die für das System im Ganzen gelten. Schneider nennt sie im Kontext von Qualitätsmodellen „Allgemeine Qualitätsziele“[32].

Unter den *Business Goals* werden in der Literatur über die schon genannten Qualitätsaspekte hinaus abstraktere Ziele genannt [vgl. z. B. 75, 17, 78]. *System Acceptance* beschreibt dabei das Ziel, die Wahrscheinlichkeit zu erhöhen, dass die Entscheidungen eines Systems von *End Usern* akzeptiert werden [75]. Auch kann ein *Business Goal* sein, die Anzahl der Nutzungen des Systems zu erhöhen. Dies kann sowohl pro Nutzer als auch insgesamt gesehen werden und ist je nach

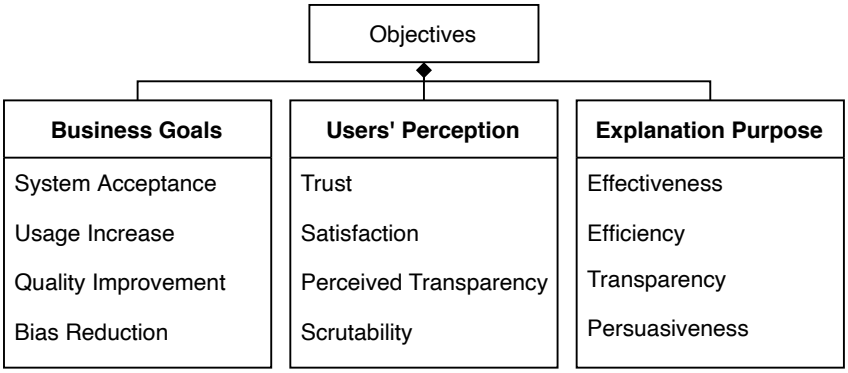


Abbildung 5.3: Übersicht über die *Objectives* eines erklärbaren Systems

Aspekt	Synonym	Quellen
Business Goals	Stakeholder Goals	[17]
	(Intended) Purpose	[69]
	Higher-level Goals	[17]
	Application Level	[37]
Users' Perception	User Perceived Quality Factors	[17]
	(Consumer) Needs	[66] [7]
	User Goals	[66]
	Intermediate Requirements	[69]
	Human Level	[37]
Explanation Purpose	(Explanation) Purpose	[17]
	Explanatory Goal	[43] [22]
	Function Level	[37]

Tabelle 5.4: Abstraktionsebenen der Zielsetzung bei der Integration von Erklärungen in ein System

Context verschieden definiert [17]. Auch kann ein sehr allgemeines Ziel sein, die grundsätzliche Qualität des Systems zu erhöhen [32].

User Perception Goals *User Perception Goals* sind jene Ziele, die direkt durch den *End User* des Systems wahrgenommen werden sollen. Sie tragen zur Erreichung der allgemeinen Ziele auf der höheren Ebene bei. Diese Ziele sind als Zwischenziele hin zu einem konkreten Ziel für zu integrierende Erklärungen zu verstehen.

Explanation Goals Die *Explanation Goals* sind „konkrete Qualitätsziele" für Erklärungen ([vgl. 32]). Diese sind allerdings keine konkreten Ziele für eine bestimmte Situation, in der die Erklärungen eingesetzt werden sollen. Hierfür müssen diese bis zu ihrer Messbarkeit weiter verfeinert werden, um daraus Anforderungen zu entwickeln (siehe Unterabschnitt 5.2.4).

Die hier vorgestellten Ziele bei der Integration von Erklärungen stellen keine abgeschlossene oder vollständige Liste dar. Sie dienen im Rahmen des Leitfadens als Überblick über in der Literatur häufig betrachtete Ziele, für deren Erreichung Erklärungen erfolgreich eingesetzt wurden.

RQ1 Welche Rahmenbedingungen haben einen Einfluss auf die Anforderungen für Erklärungen?

Mit dem *Context* des Systems und den *Objectives* für die Integration von Erklärungen wurden im Rahmen des ersten Modellteils die Rahmenbedingungen mit einem direkten Einfluss auf Anforderungen an Erklärungen vorgestellt. Die unter *External Dependencies* zusammengefassten Aspekte und Ausprägungen sind folglich die Antwort auf die erste Forschungsfrage.

Als Hilfsmittel kann der erste Teil des Modells vordergründig bei der Anforderungserhebung (*Requirements Elicitation*) und -Analyse (Requirements Analysis) helfen [32]. Auf dieser Basis können Anforderungen formuliert und die Grundlagen für Hypothesen gelegt werden. Folglich wird durch den ersten Modellteil auch die erste Leitfadenanforderung erfüllt ([GR1]).

Im folgenden Abschnitt werden jene Eigenschaften von Erklärungen betrachtet, welche in vorangegangenen Arbeiten bereits zur Umsetzung von Anforderungen verwendet wurden.

5.2.3 Eigenschaften

Die *Characteristics* umfassen jene Eigenschaften von Erklärungen, die einen Einfluss auf die externe Qualität eben dieser haben. Unter dem Aspekt sind die Möglichkeiten zusammengefasst, die bei der Ausgestaltung von Erklärungen bestehen. Gegliedert sind die Eigenschaften in den Bedarf der Erklärung (*Demand*), die ausgelieferten Informationen (*Content*) und die Bereitstellung (*Presentation*). Diese drei Unterkategorien werden in den folgenden Abschnitten vorgestellt. Wichtig zu beachten ist, dass sich die vorgestellten Möglichkeiten nicht unbedingt ausschließen. Das heißt vor allem, dass nicht jede Erklärung genau eine Ausprägung eines Aspektes erfüllen muss, sondern auch neue oder zwischen zwei Möglichkeiten liegende Eigenschaften aufweisen kann. Darüber hinaus muss auch nicht zwangsweise für jede Eigenschaftskategorie eine Eigenschaft ausgewählt werden, da nicht alle Kategorien auf jeden Kontext zutreffen.

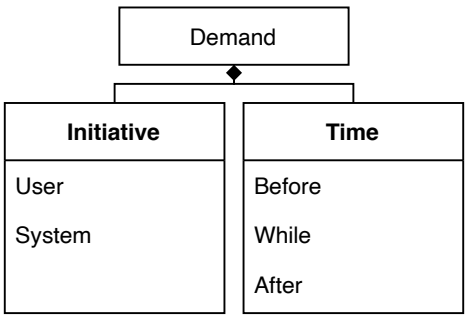


Abbildung 5.4: Übersicht über den *Demand* für Erklärungen

Bedarf

Der *Demand* für Erklärungen kann aus verschiedenen Perspektiven betrachtet werden. Bei welcher Aufgabe und bei welchen Ereignissen Erklärungen überhaupt benötigt werden, muss in den Anforderungen für Erklärungen festgehalten werden. Diese entstehen auf Basis der *External Dependencies*. Unter der Kategorie *Demand* in diesem Modell ist zusätzlich festgehalten, auf welche Initiative hin (*Initiative*) und zu welchem Zeitpunkt in Bezug auf ein Ereignis im System dem *End User* Erklärungen vom System zur Verfügung gestellt werden sollen. Einen Überblick über die Verwendung verschiedener Ausprägungen der Aspekte ist in Abbildung 5.4 und Tabelle 5.5 dargestellt.

Aspekt	Ausprägung	Quellen
Initiative	Manual	[7] [43] [70]
	Automatic	[7] [45] [70] [48] [50]
Time	Before	[30] [70] [35] [61] [52] [52]
	while	[30] [70] [35]
	After	[30] [70] [35] [61] [52] [36] [52]

Tabelle 5.5: Der Bedarf einer Erklärung zusammen mit in der Literatur untersuchten Einflüssen auf die Qualität von Erklärungen

Initiative Die *Initiative* einer Erklärung ist der Auslöser für das Geben von Erklärungen in einem System. Eine Möglichkeit ist eine automatische Auslieferung der Erklärung an den *End User*. Das System trifft dann allein die Entscheidung, wann und ob *End User* eine Erklärung bekommt (*Automatic*). Alternativ können Erklärungen vom *End User* manuell angefordert werden (*Manual*).

Time Die *Time* ist der Zeitpunkt im Verhältnis zu einem Ereignis, zu dem das System eine Erklärung bereitstellt. Rosenfeld und Richardson sowie Wiegand et al. haben explizit untersucht, wann Erklärungen angezeigt werden sollten, wenn ein Ereignis im System auftritt oder das System eine Aktion durchführt. Dies umfasst die Möglichkeiten vor dem Ereignis (*Before*), während (*While*) oder nach dem Ereignis (*After*) eine Erklärung zu diesem zu liefern [30, 70]. Letzteres wird in der Literatur zum Teil auch als *Posthoc-Explanation* referenziert [37].

Im Rahmen von **RQ2** kann an dieser Stelle der *Demand* als relevante Eigenschaft von Erklärungen für die Erklärungsqualität festgehalten werden.

Inhalt

Unter dem Punkt *Content* wird definiert, mit welchen Inhalten die *End User* durch Erklärungen versorgt werden [17]. Dieser ist einer von zwei Teilen des Modells, welcher sich auf die Granularität von Erklärungen bezieht. Der *Content* beinhaltet nicht nur den Informationstyp (*Information Type*) den das System vermittelt, sondern auch wie viel Inhalt (*Information Density*). Ein weiterer Aspekt

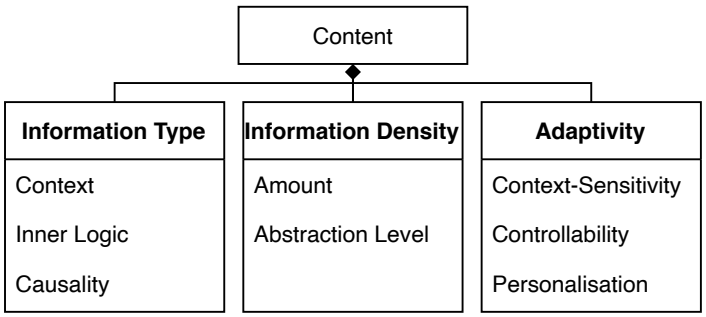


Abbildung 5.5: Übersicht über den *Content* von Erklärungen

ist Anpassungsfähigkeit der Inhalte (*Adaptivity*). Tabelle 5.6 und Abbildung 5.5 enthalten die verschiedenen Ausprägungen dieser Aspekte zusammen mit deren Anwendungen in der Literatur, welche einen Einfluss auf die externe Qualität von Erklärungen haben.

Information Type Der *Information Type* beschreibt die Inhalte, die *End Usern* mithilfe der Erklärung übermittelt werden. Unter diesem Aspekt sind in der Literatur sehr verschiedene Ansätze zu finden, die unterschiedliche Typen definieren. Beispielsweise stellen Chazette, Karras und Schneider mithilfe von Fragewörtern verschiedene Informationstypen dar [7], während Rosenfeld und Richardson selbige Fragewörter nutzt, um andere Inhalte zu beschreiben und weitere ergänzt. Grundsätzlich kann zwischen globalen Erklärungen, die immer gültig sind und situationsabhängigen (lokalen) Erklärungen unterschieden werden [80]. Zusammen mit weiteren Arbeiten [72, 68] wurden die verschiedenen Definitionen in drei verschiedene Informationstypen gebündelt. Diese fassen die in der Literatur am häufigsten transportierten Informationen zusammen. Allerdings bilden die drei Ausprägungen nicht alle möglichen Informationen in Erklärungen ab.

Aspekt	Ausprägung	Quellen
Information Type	Context	[2] [29] [74] [29] [53] [77] [17] [78]
	Inner Logic	[2] [51] [26] [77] [57] [78] [74]
	Causality	[2] [68] [50] [51] [29] [29] [53] [74] [26] [77] [57] [17][60] [78] [80] [72]
Information Density	Amount	[78] [18] [47] [65]
	Abstraction Level	[26] [47]
Adaptivity	Context-Sensitivity	[72] [74]
	Controllability	[68] [63]
	Personalization	[72] [74] [64] [43] [37]

Tabelle 5.6: Eigenschaften einer Erklärung bezogen auf den Inhalt einer Erklärung mit in der Literatur gezeigtem Einfluss auf die externe Qualität von Erklärungen

Kontextinformationen in einer Erklärung geben Auskunft über die zugrundeliegenden Daten (*Context*). Dabei werden die eingehenden Informationen auf Basis derer das System Entscheidungen trifft, dem *End User* dargelegt.

Eine weitere Möglichkeit ist das Erklären der Funktionsweise von Algorithmen eines Systems (*Inner Logic*). Dies sind die Informationen, wie genau ein System die ihm zur Verfügung stehenden Daten verarbeitet und interpretiert.

Ein dritter Weg ist die Erklärung von Zusammenhängen zwischen den Eingaben und Ausgaben des Systems (*Causality*). In einer solchen Erklärung wird den *End Usern* der Grund für ein bestimmtes Systemverhalten oder eine Entscheidung erklärt. Hierbei gibt es verschiedene Optionen, Gründe zu erläutern. Eine Erklärung dieser Art kann die Information enthalten, warum ein bestimmtes Systemverhalten in einer Situation erfolgt ist. Auch kann eine Erklärung vermitteln, warum ein alternatives Verhalten oder eine alternative Ausgabe des Systems nicht erfolgt ist [55].

Information Density Die *Information Density* beschreibt die Menge und die Kompaktheit an Informationen, die eine Erklärung enthält. Dabei ist einerseits wichtig, ob *End Usern* alle vorliegenden Erklärungsmöglichkeiten vom System angezeigt werden (*Amount*). Andererseits spielt es eine Rolle mit welchem Detailgrad die Informationen dargestellt werden (*Abstraction Level*). Zum Beispiel können *End Usern* wenig Informationen angezeigt werden, die nicht die vollen Details abdecken, um diese nicht zu überfordern.

Adaptivity *Adaptivity* definiert, wie statisch die Erklärungen in einem System sind. Eine Ausprägung ist dabei der Grad, zu dem eine Erklärung auf den aktuellen *Context* angepasst ist (*Context-Sensitivity*). Außerdem beinhaltet *Adaptivity* die *Personalisation*, welche darstellt, inwiefern Erklärungen auf den aktuellen *End User* anpassbar sind, zum Beispiel an dessen Expertise. *Controllability* beschreibt dabei, welchen Einfluss *End User* haben, mit der Erklärung zu interagieren. Unter diesen Aspekt fällt unter anderem die Möglichkeit, dass *End User* Erklärungen optional anfordern können oder sie innerhalb eines Erklärungsdialogs navigieren können.

An dieser Stelle kann für **RQ2** hinzugefügt werden, dass auch der *Content* einer Erklärung einen großen Einfluss auf die externe Qualität von Erklärungen hat. Dies lässt sich unter anderem an der Anzahl an Autoren festmachen, die verschiedene Facetten des Einflusses durch den *Content* von Erklärungen auf die Qualität untersuchen.

Presentation

Nachdem nun sowohl der *Demand* als auch der an den *End User* übermittelte Inhalt als zentrale *Characteristics* von Erklärungen vorgestellt wurden, fehlt im Modell die Art der Präsentation der Erklärung an den *End User*. Diese ist mit ihren zugehörigen Ausprägungen unter *Presentation* zusammengefasst (siehe Abbildung 5.6). Sie stellt den zweiten Teil der Granularität von Erklärungen dar. Zu dem Aspekt gehören in diesem Modell für Erklärungen das Medium (*Medium*), über das die Erklärung *End Usern* bereitgestellt wird, der verwendete Ton (*Tone*) des *Explainers* (siehe

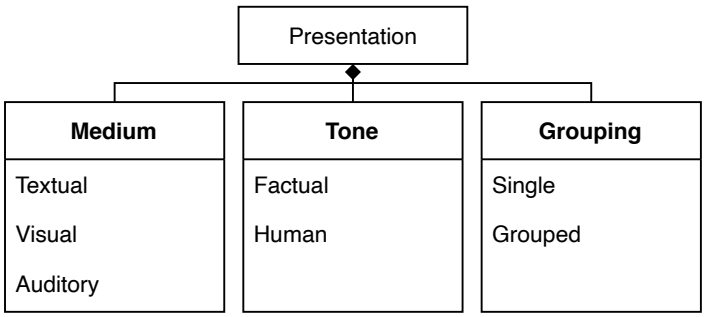


Abbildung 5.6: Übersicht über die *Presentation* von Erklärungen

Unterabschnitt 2.1.1) [vgl. 5] und die Gruppierung von Erklärungen respektive Erklärungstypen (*Grouping*). Tabelle 5.7 stellt die Ausprägungen zusammen mit der Literatur, die den entsprechenden Aspekt in Bezug auf die externe Qualität von Erklärungen untersucht, dar.

Medium Das *Medium* einer Erklärung ist der Informationsträger für die *Presentation* der Inhalte. Dabei können verschiedene Möglichkeiten aus dem Multimedia-Bereich verwendet werden. In der Literatur untersucht wurden Texterklärungen (*Textual*), visuelle Darstellungen wie z. B. Diagramme (*Visual*) sowie auditive Erklärungen (*Auditory*). Insbesondere dieser Aspekt bietet Möglichkeiten für Mischformen und Kombinationen [18].

Tone Der *Tone* einer Erklärung bestimmt die Art, wie *End Usern* der Inhalt einer Erklärung näher gebracht wird. Das Spektrum der Möglichkeiten erstreckt sich dabei vor allem zwischen sehr faktisch gehaltenen Erklärungen (*Factual*) und persönlichen bzw. menschlichen Erklärungen (*Human*). Ein Beispiel wäre die Bereitstellung von Erklärungen über einen persönlichen Assistenten [48].

Grouping Mit dem *Grouping* von Erklärungen wird bestimmt, wie viele Erklärungen *End Usern* gleichzeitig beziehungsweise kombiniert präsentiert werden. Zum Beispiel können erklärende Grafiken wie Graphen mit Texterklärungen kombiniert werden. Grundsätzlich gibt es dabei die Möglichkeiten eine Erklärung zu einem Zeitpunkt anzuzeigen (*Single*) oder mehrere Erklärungen bzw. Erklärungstypen zu gruppieren (*Grouped*). Bei letzterem können entweder Erklärungen mit verschiedener *Presentation* aber dem gleichen *Content* kombiniert werden [18] oder mehrere einzelne Erklärungen zusammen dargestellt werden [22].

Als letzter Teil der Antwort auf **RQ2** kann die *Presentation* mit den dazugehörigen Aspekten als Eigenschaft von Erklärungen mit einem Einfluss auf die externe Qualität von Erklärungen ergänzt werden.

Aspekt	Ausprägung	Quellen
Medium	Textual	[37] [22] [43] [51] [45] [45] [68] [74] [17]
	Visual	[37] [51] [67] [68] [17] [58]
	Auditory	[36] [17] [59]
Tone	Factual	[45] [68] [35] [57]
	Human	[68] [35] [49] [29] [57]
Grouping	Single	[17] [22] [51] [45] [68]
	Grouped	[17] [22] [43]

Tabelle 5.7: Verschiedene Übermittlungsmöglichkeiten für Erklärungen an den *End User*, die in der Literatur einen Effekt auf die externe Qualität von Erklärungen gezeigt haben

RQ2 Welche Eigenschaften von Erklärungen haben einen Einfluss auf die externe Qualität eines erklärbaren Systems?

Der in diesem Abschnitt vorgestellte Teil des Modells für Erklärungen (*Characteristics*) beinhaltet die Antwort auf die zweite Forschungsfrage. Dabei sind explizit die Eigenschaften *Demand*, *Content* und *Presentation* mit einem Einfluss auf die externe Qualität von Erklärungen zu benennen. Die genauen Ausprägungen dieser Eigenschaften werden als Unterpunkte der jeweiligen Aspekte im Modell enthalten. Die Ergebnisse, die in dem Modell zusammengefasst sind, entspringen dabei der Evaluation von Erklärungen mit verschiedenen Eigenschaften in der Literatur.

5.2.4 Evaluation

In diesem Abschnitt des Modells wird die Operationalisierung der externen Qualität von Erklärungen erläutert. Dabei werden sowohl die Eigenschaften von integrierten Erklärungen an sich beschrieben als auch Effekte, die durch den Einfluss von Erklärbarkeit auf weitere Qualitätsaspekte messbar sind. Dabei werden die schon in Unterabschnitt 5.2.2 vorgestellten Qualitätsziele betrachtet.

Es wird von mehreren Autoren festgestellt, dass das Messen der Qualität von Erklärungen aufgrund der stark unterschiedlichen subjektiven Wahrnehmung von *End Usern* eine Herausforderung darstellt [17, 45, 18]. Daher werden in vorhandener Literatur zur Evaluation von Erklärungen vorwiegend subjektive Fragebögen verwendet. Diese orientieren sich in der Regel an den zuvor ausgewählten Zielen für zu integrierende Erklärungen (siehe Unterabschnitt 5.2.2).

Grundsätzlich wird bei der Evaluation zwischen subjektiven Metriken (*qualitative* [34]) und Verhaltensmetriken (*quantitative* [34]) unterschieden. Entsprechend sind die verschiedenen Evaluationsmethoden im Modell in *Qualitative Research* und *Quantitative Research* gegliedert. Waa et al. betonen, dass zur Beurteilung der Qualität von Erklärungen beide Methoden zusammen eingesetzt werden sollten. Sie schreiben, dass es so möglich ist, die „komplette Perspektive“ auf eine Erklärung zu erhalten [übersetzt vgl. 69]. Eine Übersicht der verschiedenen Möglichkeiten ist

in Abbildung 5.7 abgebildet.

Zudem werden in der Literatur einerseits Studien durchgeführt, bei denen die Teilnehmer jeweils mehrere Bedingungen, wie z. B. verschiedene Erklärungstypen nacheinander evaluieren (*with-in subject*). Allerdings finden auch Studien Anwendung, bei denen pro Teilnehmer nur eine Bedingung evaluiert und diese dann verglichen werden (*between subject*).

Des Weiteren stehen grundsätzlich die von Wohlin et al. definierten Evaluationsstrategien, die bereits in den Ergebnissen der Literaturrecherche (Kapitel 4) vorgestellt wurden, zum Messen der Qualität von Erklärungen zur Verfügung: *Survey*, *Case Study* und *Experiment*. Diese werden unter anderem von Ribera und Lapedriza sowie Doshi-Velez und Kim auf Erklärbarkeit übertragen. Die verwendete Strategie hängt dabei vom *Context* und den *Objectives* ab. Je nachdem, welche Ergebnisse von Stakeholdern benötigt werden, muss die Evaluation kontrollierter oder weniger kontrolliert sein. Die Vor- und Nachteile dessen sind bei Wohlin et al. nachzulesen [34].

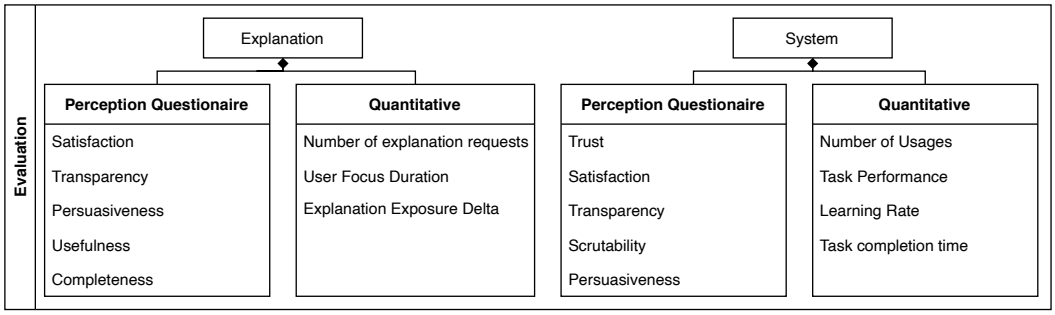


Abbildung 5.7: Übersicht über die *Evaluation* von Erklärungen

Qualitative Evaluation

Bei der Durchführung von qualitativen Evaluationen von Erklärungen gibt es verschiedene Möglichkeiten. Einerseits kann die Qualität der Erklärungen direkt gemessen werden. Alternativ wurden in der Literatur aber auch die bereits vorgestellten Qualitätsaspekte (siehe Unterabschnitt 5.2.2) verwendet, um über den Einfluss von Erklärungen auf diese, die externe Qualität von Erklärungen abzuleiten.

Neben den in Unterabschnitt 5.2.2 vorgestellten Qualitätszielen, gibt es weitere Aspekte, die in der Literatur zur Messung der direkten Qualität von Erklärungen vorgestellt wurden [51]. Diese werden im Folgenden erläutert.

Usefulness *Usefulness* oder auch *Helpfulnesss* ist der Grad zu dem *End User* das subjektive Empfinden haben, dass eine Erklärung sie bei der Nutzung oder dem Verständnis über ein System unterstützt hat.

Completeness *Completeness* ist das subjektive Empfinden von *End Usern*, dass gegebene Erklärungen einen vollständigen Aufschluss über den erklärten Systembestandteil geben und keine Informationen weglassen.

Bei der subjektiven Evaluation wird in der Literatur vorwiegend die Bewertung verschiedener Aussagen durch Studienteilnehmer eingesetzt [51, 44, 59]. In der Regel ist dabei die Zustimmung oder Ablehnung einer Aussage von Interesse [84, 62, 61, 47]. Tabelle 5.8 und Tabelle 5.9 stellen eine Übersicht von verwendeten Aussagen für die Messung der Qualitätsaspekte dar, über welche die Erklärungsqualität messbar ist. Zusammengefasst sind hierbei nur verallgemeinerbare Aussagen und nicht solche, die nur einen spezifischen *Context* betreffen.

In spitzen Klammer sind Platzhalter dargestellt, um die aufgelisteten Aussagen besser auf einen *Context* anpassen zu können. „<System>“ steht entweder für das aktuelle System, Teilsysteme, Tools oder Algorithmen, welche evaluiert werden sollen. „<Erklärung>“ steht für eine spezifische Erklärung, die evaluiert wird. Die Aufgabe(n), die ein *End User* während einer Evaluation erledigt werden, können im Platzhalter „Aufgabe“ eingesetzt werden. Darüber werden in eckigen Klammern optionale Teile von Aussagen angegeben, die nur für manche Evaluationskontexte sinnvoll sind.

Qualitätsaspekt	Aussage
Satisfaction	<Erklärung>, stellt mich mit meinem Verständnis über <System> zufrieden. [vgl. 54] <Erklärung>, ist zufriedenstellend. [vgl. 54, 84, 22]
Perceived Transparency	Die Informationen der Erklärung waren ausreichend, um <Aufgabe> gut zu erfüllen. [vgl. 59, 22]
Persuasiveness	<Erklärung>, ist überzeugend. [vgl. 51, 44] <Erklärung>, weckt Interesse. [vgl. 51, 44]
Usefulness	<Erklärung>, ist einfach zu verstehen. [vgl. 51, 44] <Erklärung>, ist nützlich bei der Erfüllung von <Aufgabe>. [vgl. 51, 44, 84, 22]
Completeness	<Erklärung>, ist hinreichend vollständig. [vgl. 84, 54] <Erklärung>, ist hinreichend detailliert. [vgl. 54]

Tabelle 5.8: Aussagen zur qualitativen Evaluation ausgewählter externer Qualitätsaspekte in Bezug auf Erklärungen in einem System

In Tabelle 5.8 werden die Aussagen zur Evaluation zusammengefasst, die nur für *End User* relevant sind, die eine Erklärung direkt evaluieren sollen. Die Aussagen beziehen sich dabei direkt auf die Eigenschaften der Erklärung.

In Tabelle 5.9 werden die Aussagen zusammengefasst, welche auf beide der zuvor erwähnten Situationen anpassbar sind. Je nachdem, ob die aktuelle Messung sich direkt auf eine Erklärung bezieht, müssen die optionalen Teile dabei weggelassen werden. Wenn alle optionalen Teile eingefügt werden, kann ein Vergleich zwischen verschiedenen Bedingungen gezogen werden. Die Nutzerpräferenz ist bei *within-subject* Studien möglich.

Die Aussagen aus Tabelle 5.9 können darüber hinaus zum Teil auch als Entschei-

Qualitätsaspekt	Aussage
Trust	Ich kann [mithilfe von <Erklärung>] [besser] sagen, wie vertrauenswürdig das System ist. [vgl. 84, 22, 49, 47]
Satisfaction	Ich bin mit der Nutzung von <System> zufrieden[er, mithilfe von <Erklärung>]. [vgl. 22] Ich werde <System> [wegen <Erklärung>] wieder benutzen. [vgl. 22]
Transparency	Ich kann den Entscheidungsprozess von <System> [mithilfe von <Erklärung>] [besser] verstehen. [vgl. 59, 22] Ich kann [mithilfe von <Erklärung>] [besser] verstehen, wie <System> funktioniert. [vgl. 54, 84, 47]
Scrutability	Ich kann [mithilfe von <Erklärung>] [besser] sagen, wie zuverlässig <System> ist. [vgl. 84, 22]
Persuasiveness	Ich bin von den Entscheidungen von <System> [mithilfe von <Erklärung>] überzeugt. [vgl. 56]

Tabelle 5.9: Aussagen zur qualitativen Evaluation ausgewählter Qualitätsaspekte in Bezug auf ein System oder Systemteile

dungsfragen formuliert werden. Wenn dies der Fall ist, wurde in der Literatur für eine der beiden Antwortmöglichkeiten zusätzlich ein Freitextfeld für eine Begründung bereitgestellt, z. B. „Hätten Sie sich gewünscht, dass die Erklärungen zusätzliche Informationen enthalten? Wenn ja, welche Art von Informationen und wann, d. h. in welchen Situationen?“[übersetzt vgl. 54].

Als Alternative zu den vorgestellten Aussagen haben außerdem einige Autoren direkt die Präferenz zwischen verschiedenen Formen von Erklärungen abgefragt, statt diese indirekt über Aussagen zu ermitteln [18, 67, 68, 22, 70, 71, 72].

Des Weiteren gibt es mehrere Beispiele dafür, dass Autoren bei der Messung der Qualität auch den *Mental Workload* von *End Usern* bei der Erfüllung von Aufgaben gemessen haben. Dabei wird überprüft, wie sehr die kognitiven Fähigkeiten der Probanden gefordert werden. Alle betrachteten Beispiele aus der Literatur verwenden dabei den standardisierten Fragebogen „NASA TLX“ [36, 70, 73].

Über die bereits vorgestellten Möglichkeiten zur qualitativen Evaluation von Erklärungen untersuchen außerdem mehrere Autoren den Einfluss von Erklärungen auf *Usability*. *Usability* ist dabei definiert als der „Grad, in dem ein Produkt oder System von bestimmten Nutzern verwendet werden kann, um bestimmte Ziele mit *Efficiency*, *Effectivity* und *Satisfaction* in einem bestimmten *Context* zu erreichen “ [übersetzt vgl. 6]. Die Beispiele aus dem Bereich der Erklärbarkeit fokussieren sich dabei auf die Verwendung von *Think-Aloud*-Studien, bei denen die Studienteilnehmer während der Nutzung eines Systems ihre Gedanken laut

aussprechen und sich so *Usability*-Probleme erfassen lassen [70, 50]. Hintergrund ist, dass *Explainability* auch negative Effekte auf *Usability* haben kann [5]. Diese Effekte sollten daher auch gemessen werden.

Die zuvor erläuterten qualitativen Metriken zur Messung der Qualität von Erklärungen sind der erste Teil der Antwort dieses Modells auf **RQ3**. Dabei kann festgehalten werden, dass der qualitative Teil der Evaluationen von Erklärungen den Großteil der in der Literatur bereits erfolgten Evaluationsmethoden ausmacht und dort aufgrund der stark subjektiven Wahrnehmung von Erklärungen als notwendiger Teil erachtet wird.

Quantitative Evaluation von Erklärungen

Die quantitative Messung der Qualität von Erklärungen wird in der Literatur seltener verwendet, als die qualitative. Dies liegt unter anderem an den unterschiedlichen subjektiven Meinungen von *End Usern* in Bezug auf den Bedarf für Erklärungen und dessen Granularität [7, 18]. Dennoch wird in der Literatur empfohlen, zusätzlich auch eine quantitative Evaluation durchzuführen [22]. Beispielsweise kommen Wiegand et al. zu dem Ergebnis, dass *End User* subjektiv andere Erklärungen präferieren können, als zu einer besseren Performanz der Aufgabenerfüllung der *End User* führen würden [36].

Bei direkter Messung von quantitativen Aspekten von Erklärungen werden in der Literatur neben Domänen- oder Erklärungs-spezifischen Metriken, solche eingesetzt, die unmittelbar den Bedarf und die Granularität messen. In Tabelle 5.10 sind die Metriken aufgeführt, welche Kontext- und Erklärungsunabhängig verallgemeinerbar sind.

Des Weiteren kann gemessen werden, wie häufig *End User* eine Erklärung anfordern, sofern diese interaktiv sind (*Number of explanation requests*). Daraus kann abgeleitet werden, wie groß der Bedarf für eine Erklärung in der spezifischen Situation wirklich ist [70].

Aspekt	Metrik	Quellen
Bedarf	*Number of explanation requests*	[70] [7] [22]
Granularität	*User Focus Duration*	[22]
	Explanation Exposure Delta	[17]

Tabelle 5.10: Quantitative Evaluationsmöglichkeiten zur direkten Messung der Qualität von Erklärungen

Zusätzlich kann nach der Anforderung einer Erklärung durch einen *End User* gemessen werden, wie lange dieser sich auf diese fokussiert (*User Focus Duration*) [22]. Daraus kann abgeleitet werden, ob die Granularität der Erklärung für den *End User* passend war. Eine weitere Möglichkeit, die Granularität zu messen ist außerdem das *Explanation Exposure Delta* [17]. Für das *Explanation Exposure Delta* wird eine Metrik zum Verständnis des Systems benötigt, welche auf eine Intervallskala abbildet. Die verwendeten Metriken sind dabei sehr Domänen- und Aufgaben-spezifisch. Diese Metrik wird im Anschluss einmal vor und einmal nach

der Anzeige einer Erklärung gemessen. Die Differenz dieser Werte ist das *Explanation Exposure Delta*. Umso größer das Exposure Delta ist, umso mehr Wissen haben *End User* durch eine Erklärung erlangt.

Neben der direkten Evaluation der Eigenschaften von Erklärungen haben deutlich mehr Autoren die Auswirkungen von Erklärungen auf bestimmte Qualitätsaspekte von Systemen untersucht. Carvalho et al. stellen in ihrer Veröffentlichung eine Übersicht über verschiedene Metriken vor, um u. a. die in dieser Arbeit erwähnten Qualitätsaspekte zu messen [42] (siehe Unterabschnitt 5.2.2).

Qualitätsaspekt	Metrik	Quellen
Trust	*Number of Usages*	[81]
Persuasiveness	*Task Performance*	[81]
Effectiveness	*Task Performance*	[69] [67] [43] [68] [53] [65] [55] [79] [35] [81]
	Learning Rate	[43] [79]
Efficiency	*Task Completion Time*	[81]

Tabelle 5.11: Evaluation

In Tabelle 5.11 werden die in der Literatur in Bezug auf Erkärbarkeit genutzten Metriken zusammengefasst. Für den Qualitätsaspekt *Trust* ist die Anzahl der Nutzungen (*Number of Usages*) aufgeführt. Diese kann sowohl absolut als auch relativ zu einem Zeitraum betrachtet werden. Die *Number of Usages* kann außerdem in unterschiedlichen Anwendungsfällen verschieden definiert werden. Beispielsweise kann sie sich auf die Anzahl der Nutzungen einer Software im Allgemeinen oder einer bestimmten Funktion beziehen. Die *Task Performance* ist hier als abstrakte Metrik aufgeführt. Dieser Begriff wird von mehreren Autoren als Qualität der Aufgabenausführung von *End Usern* definiert. Konkrete Metriken werden in dieser Arbeit allerdings nicht aufgeführt, da diese laut mehrerer Autoren sehr *Context*-abhängig sind [81, 79]. Die *Task Performance* wird dabei sowohl für die Bewertung der *Persuasiveness* des Systems als auch die *Efficiency* herangezogen. Für letztere haben Tintarev und Masthoff sowie Gunning und Aha außerdem die *Learning Rate* als Evaluationsfaktor herangezogen. Sie vergleichen dabei unter verschiedenen Bedingungen, wie schnell sich die *Task Performance* von *End Usern* erhöht.

Für die Messung der genannten Qualitätsaspekte stellen die vorgestellten Metriken keine vollständige Liste dar. Sie sollen lediglich einen Überblick über häufig verwendbare Möglichkeiten zur Evaluation geben.

RQ3 Auf welche Art und Weise kann evaluiert werden, ob die in ein erklärbares System integrierten Erklärungen das Ziel der Integration bezogen auf externe Qualitätsaspekte erfüllt haben?

Der Abschnitt *Evaluation* des Modells für Erklärungen ist die Antwort auf die dritte Forschungsfrage. Dies erfolgt in drei Teilen. Zunächst enthält das Modell einen kurzen Überblick über das Studiendesign für verschiedene Anwendungsfälle. Außerdem sind Teil der Antwort die zwei verschiedene Evaluationsmöglichkeiten: direkte Erklärungsevaluation und Evaluation der Auswirkungen auf das System im Allgemeinen. Für beide Varianten werden in dem Modell Beispiele für qualitative und quantitative Betrachtungen genannt. Dies erfüllt auch die Anforderung an den Leitfaden ([GR3])

Zusammenfassend kann gesagt werden, dass für ein ganzheitliches Bild über die Qualität von Erklärungen sowohl qualitativ als auch quantitativ die Erklärungen an sich sowie dessen Auswirkung auf weitere externe Qualitätsaspekte betrachtet werden müssen [22]. Wichtig ist dabei auch die Betrachtung von Aspekten zur Kontrolle, auf die ggf. negative Effekte durch Erklärungen möglich sind.

Nachdem durch das in den vorigen Abschnitten vorgestellte Modell die Forschungsfragen **RQ1** - **RQ3** beantwortet wurden, erfolgt im Weiteren eine Zusammenfassung der Abhängigkeiten der einzelnen Aspekte des Modells. Neben der Beantwortung der Forschungsfragen **RQ4.1** und **RQ4.2** ist dies relevant, um festzulegen, welche Qualitätsaspekte bei der Evaluation von Erklärungen betrachtet werden müssen.

5.3 Abhängigkeiten

Zusätzlich zum Modell über die in der Literatur bereits betrachteten Aspekte von Erklärungen sind im Folgenden die Ergebnisse der Auswirkungen dieser Aspekte auf die externe Qualität von Erklärungen zusammengefasst. Die aufgelisteten Einflüsse sind im Rahmen der Literaturrecherche erarbeitet worden. Die Quellen der Zusammenhänge werden in der folgenden Übersicht jeweils mit angegeben.

Enthalten sind jene Resultate, die entweder bereits bewiesen sind oder die Autoren eine starke Vermutung für eine These haben, die allerdings noch einer Überprüfung bedarf. Noch nicht verifizierte Resultate sind mit (*) gekennzeichnet. Auch sind nicht alle Ergebnisse vorangegangener Arbeiten *Context*-unabhängig gezeigt worden. Somit können die Ergebnisse zum Teil nur mit Einschränkungen auf andere Kontexte übertragen werden. Außerdem sind manche Einflüsse nur in Vergleich zu alternativen Erklärungen gezeigt worden. Auch diese Einschränkungen werden in der Auflistung mit aufgeführt.

Generell ist einerseits das Ziel dieser Zusammenfassung von Abhängigkeiten, den Anwendern dieses Leitfadens eine Unterstützung bei der Ableitung von konkreten Anforderungen an Erklärungen aus den *Explanation Purposes* zu geben. Ebenso soll dieser Überblick über zuvor geleistete Ergebnisse auch bei der Umsetzung der Anforderungen im System unterstützen. Folglich dient dieser Teil des Leitfadens der Beantwortung der Forschungsfragen **RQ4.1** und **RQ4.2**, welche explizit die Zusammenhänge zwischen verschiedenen Aspekten des Modells abdecken. Damit

gibt der Leitfaden zusätzliche Informationen zu den Einflüssen, die einzelne Eigenschaften von Erklärungen unter bestimmten Rahmenbedingungen auf externe Qualitätsaspekte haben. Je nach *Context* können dabei auch Konflikte zwischen den externen Qualitätsaspekten, auf die Erklärbarkeit Auswirkungen hat, entstehen. Die Erkennung dieser soll durch den folgenden Katalog der Zusammenhänge unterstützt werden.

Wie schon das Modell der Aspekte von Erklärungen hat auch dieser Katalog mit den Abhängigkeiten keinen Anspruch auf Vollständigkeit, sondern soll einen guten Überblick über bestehende Ergebnisse schaffen, die sich auf viele Anwendungsfälle übertragen lassen.

Die Formulierung der Abhängigkeiten orientiert sich an der Formulierungsweise von Carvalho, Andrade und Oliveira bei der generellen Untersuchung verschiedener NFRs [1]. Das Konzept beschreibt, dass einzelne Eigenschaften von Softwaresystemen einen positiven, negativen oder neutralen Effekt auf Qualitätsaspekte haben können. Bedarf ein Zusammenhang einer Bedingung, so wird diese nachgestellt.

Gegliedert ist der Katalog der Abhängigkeiten in zwei Abschnitte. Ersterer behandelt die Auswirkungen, die die Rahmenbedingungen auf den *Demand* für Erklärungen haben. Im zweiten Teil werden die Einflüsse der Rahmenbedingungen eines Systems auf die Granularität von Erklärungen, also den *Content* und die *Presentation* vorgestellt. Außerdem sind die verschiedenen Einflüsse in positive, negative und neutrale Einflüsse gegliedert.

Demand

Im Folgenden sind die Einflüsse aufgeführt, die sich darauf beziehen, ob und wann Erklärungen benötigt werden.

Positive Zusammenhänge

- Das Präsentieren von Erklärungen hat einen positiven Einfluss auf die *Perceived Transparency*, *wenn End User geringes Vertrauen in Technologie haben* [56].
- Das Präsentieren von Erklärungen hat einen positiven Einfluss auf die *Perceived Transparency*, *wenn End User Datenschutzbedenken haben* [56].
- Das Präsentieren von Erklärungen hat einen positiven Einfluss auf *Trust*, *wenn End User sich in unbekannten Situationen befinden* [52].
- Das Präsentieren einer Erklärung vor einer unbekannten Situation (*before*) hat einen positiveren Einfluss auf *Trust*, als Präsentation danach [52].
- Das proaktive Präsentieren von Erklärungen (*System-Initiative*) in unbekannten Situationen hat einen positiven Einfluss auf *Trust* [60].
- Das Präsentieren von Erklärungen für sicherheitsrelevante Systemfunktionen hat einen positiven Einfluss auf *Trust* [36].
- Das Präsentieren von Erklärungen für sicherheitsrelevante Systemfunktionen hat einen positiven Einfluss auf *Satisfaction* [36].

- Die Möglichkeit Erklärungen anzufordern *User-initiative* hat einen positiven Einfluss auf *Satisfaction* [7].

Negative Zusammenhänge

- Das Präsentieren von Erklärungen kann einen negativen Einfluss auf die *Effectivity* von *End Usern* haben, *wenn diese sehr selbstbewusst sind** [48].
- Das Präsentieren von Erklärungen kann einen negativen Einfluss auf die *Satisfaction* haben, *wenn End User hohen Trust in das System haben** [30, 25].

Weitere Zusammenhänge

- Der Zusammenhang zwischen dem Grad der Systembeeinflussung durch *End User* und dem Erklärungsbedarf ist antiproportional. Umso mehr Beeinflussungsmöglichkeiten *End User* in einem System haben, umso weniger Erklärungsbedarf haben diese [30].

RQ4.1 Welchen Einfluss hat der Bedarf von Erklärungen auf die externe Qualität eines erklärbaren Systems unter bestimmten Rahmenbedingungen?

Der erste Teil des hier gezeigten Kataloges von Zusammenhängen der Aspekte des Modells für Erklärungen beantwortet die Forschungsfrage **RQ4.1**. Er stellt dar, unter welchen Bedingungen das Präsentieren von Erklärungen, welche Einflüsse auf ausgewählte externe Softwarequalitätsaspekte hat. Die Frage wird dabei insofern beantwortet, als die Antwort aus der Literaturrecherche entstandene verallgemeinerbare Einflüsse zusammenfasst.

Granularität

Auf welche Weise der *Content* und die *Presentation* durch Rahmenbedingungen wie dem *Context* und den *Objectives* für Erklärungen beeinflusst werden, ist im Folgenden dargestellt.

Positive Zusammenhänge

- Erklärungen, welche mehrere Medien kombinieren, haben einen größeren positiven Einfluss auf die *Persuasiveness* eines Systems, als das Präsentieren der einzelnen Erklärungen [51, 35, 51, 58, 80].
- Das Präsentieren von Erklärungen, welche die zugrunde liegenden Daten für ein Systemverhalten darlegen (*Context*), haben einen positiven Einfluss auf die *Persuasiveness* eines Systems* [51, 68]. Gezeigt ist dies nur für *Recommender Systems*.
- Das Präsentieren von Erklärungen, welche die zugrunde liegenden Daten für ein Systemverhalten darlegen (*Context*), haben einen positiven Einfluss auf die *Usefulness* einer Erklärung* [51, 68]. Gezeigt ist dies nur für *Recommender Systems*.

- Das Präsentieren von kausalen Erklärungen, die den Grund angeben, warum eine Alternative nicht genommen wird, hat einen positiveren Einfluss auf die *Perceived Transparency* als das Präsentieren solcher, die den Grund angeben, warum eine Entscheidung getroffen wurde [55, 58, 57].

- Das Präsentieren von Erklärungen durch Agenten (*Human*) hat einen positiven Einfluss auf den *Trust* in Systemen* [49]. Gezeigt ist dies nur für *XAI-Systems*.

- Das Präsentieren von Erklärungen, welche die zugrunde liegenden Daten für ein Systemverhalten darlegen (*Context*), hat einen positiveren Einfluss auf die *Satisfaction* mit einem System, als das Präsentieren solcher, die die Gründe für eine Entscheidung erläutern, *wenn die End User wenig Domänenwissen auf dem Einsatzgebiet des Systems haben* [72, 55]

- Einfache Erklärungen (*Amount / Abstraction Level*) haben einen positiven Einfluss auf die *System Acceptance* [85, 23].

- Einfache Erklärungen (*Amount / Abstraction Level*) haben einen positiven Einfluss auf die *Satisfaction* mit dem System [85, 23].

- Interaktive Erklärungen haben einen positiven Einfluss auf die *Perceived Transparency* eines Systems [63].

Negative Zusammenhänge

- Das Präsentieren von Erklärungen hat einen negativen Einfluss auf den *Trust* in ein System, *wenn der Content der Erklärung nicht dazu führt, dass das Mental Model der End User danach mit der wirklich Funktionsweise des Systems übereinstimmt (Completeness)* [58, 7].

- Das Präsentieren von Erklärungen hat einen negativen Einfluss auf den *Trust* in ein System, *wenn die Erklärungen zu unseriös sind* [59].

- Das Präsentieren von Erklärungen, welche das aktuelle Systemverhalten beschreiben, kann einen negativen Effekt auf die *Satisfaction* mit dem System haben* [61].

- Das Präsentieren von Erklärungen, welche das aktuelle Systemverhalten beschreiben, kann einen negativen Effekt auf die *Effectivity* haben* [61].

- Das Präsentieren von Erklärungen kann einen so großen positiven Effekt auf *Trust* haben, dass es einen negativen Effekt auf die *Scrutability* und daher einen negativen Effekt auf die *Effectivity* der *End User* hat* [19, 79].

Neutrale Zusammenhänge

- Das Präsentieren von kausalen Erklärungen hat keinen Einfluss auf die *Effectivity* von *End Usern*, *wenn diese subjektiv unbekannt mit der Aufgabe sind - unabhängig von der objektiven Kompetenz* [48].

- Das Präsentieren von Erklärungen, die die Funktionsweise von Algorithmen im Allgemeinen erklären (*Inner Logic*), haben einen geringen Einfluss auf weitere Qualitätsaspekte [7].

- Die wirkliche *Transparency* hat keinen Einfluss auf den *Trust* der *End User* in das System, *wenn das Mental Model der End User mit der wirklichen Funktionsweise des Systems übereinstimmt* [45, 54].

- Die wirkliche *Transparency* von Erklärungen hat keinen Einfluss auf den *Trust* der *End User* in das System, *wenn die End User mit der Erklärung zufrieden sind* [45, 54].

RQ4.2 Welchen Einfluss hat die Granularität von Erklärungen auf die externe Qualität eines erklärbaren Systems unter bestimmten Rahmenbedingungen?

Die Einflüsse, die Eigenschaften der Granularität von Erklärungen auf einige externe Qualitätsaspekte haben, werden im zweiten Teil des Katalogs der Zusammenhänge der Aspekte von Erklärungen, zusammengestellt. Ebenfalls wird somit eine Antwort auf die letzte Forschungsfrage gegeben und erfüllt die Anforderung GR4, welche fordert, dass der Leitfaden eine Unterstützung bei der Auswahl von Eigenschaften bei der Umsetzung von Erklärungen geben soll.

Zusammenfassend sind mit dem Katalog der Zusammenhänge sowie dem zuvor vorgestellten Modell die Forschungsfragen beantwortet worden. Um die Anwendung des Leitfadens weiter zu vereinfachen, die Anforderungen an diesen und damit das Ziel dieser Arbeit besser zu erfüllen, werden im Folgenden konkrete Design-Auswirkungen für Erklärungen vorgestellt.

5.4 Design Auswirkungen

Dieser Abschnitt des Leitfadens zur Integration von Erklärungen in ein System enthält konkrete Empfehlungen für das Design von Erklärungen. Mit Design ist dabei nicht explizit ein visuelles Design gemeint, sondern die generelle Umsetzung.

Aus den zuvor vorgestellten Zusammenhängen werden im Folgenden Heuristiken abgeleitet, welche für die Integration von Erklärungen beachtet werden sollten. Dabei wurden die am häufigsten festgestellten und ohne Einschränkungen anwendbaren Zusammenhänge in Bedingungen für ein erklärbares System umformuliert und zusammengefasst. Die Heuristiken wurden dabei jeweils mit einem Titel versehen, ähnlich wie die *Usability heuristics for user interface design* von Nielsen [86].

10 Heuristiken für die Gestaltung von Erklärungen

1. **Accessibility** Ein erklärbares System sollte *End Usern* zu jeder Zeit die Möglichkeit geben, auf Erklärungen zugreifen zu können. [36, 7, 70, 49]. Dies umfasst auch, dass optional noch detailliertere Informationen angeboten werden sollten für *End User* mit einem größeren Bedürfnis für Informationen. [55]

2. **Context Sensitivity** Erklärungen sollten sich sehr stark am *Context* des erklärbaren Systems ausrichten. Nur so kann gewährleistet werden, dass diese *End Usern* auch helfen. [44, 76, 7].

3. **Key Information Delivery** Jede Erklärung sollte mindestens die Hauptinformation enthalten (Daten oder Begründung), die zu einem bestimmten Systemverhalten geführt hat [55]. Diese sollte grundsätzlich vollständig sein (*Completeness*) [54].

4. **Hybrid Style** Das System sollte den *End Usern* wenn möglich, den *Content* als Kombination verschiedener Erklärungsstile präsentieren [51, 35, 51, 58, 80].

5. **Minimally Complete Explanation (MCE)** [vgl. 29] Erklärungen sollten so kurz und präzise wie möglich, aber vollständig sein [70, 36, 29]. Dies gilt insbesondere bei zeitkritischen Aufgaben.

6. **Earliest viable Time** Erklärungen sollten den *End Usern* so früh wie möglich präsentiert werden. Dabei gilt der Grundsatz: Umso einfacher und kürzer eine Erklärung ist, umso früher kann sie präsentiert werden [85, 23].

7. **User Perception** Die Wahrnehmung der *End User* sollte im Zentrum von Erklärungen stehen [54].

8. **User Performance** Die Performanz der *End User* muss als Faktor bei der Integration von Erklärungen in ein System betrachtet werden [54].

9. **Education** Die, den *End Usern* gegebenen Erklärungen sollten den Lernprozess dieser beachten, um z. B. die Inhalte und Notwendigkeit von Erklärungen zu adaptieren [21].

10. **Visibility of the System Confidence** Bei Entscheidungen vom System sollte, wenn möglich sichtbar sein, wie sicher sich das System mit einer Entscheidung ist. [70, 87]

Neben den vorgestellten Heuristiken können für bestimmte Erklärungstypen außerdem weitere bestehende Heuristiken oder Empfehlungen angewendet werden. In der Regel sind Erklärungen in einem *User Interface* verortet oder können zumindest von dort aus aktiviert werden. Folglich sollten bei der Integration von Erklärungen im *User Interface* unter anderem die bereits erwähnten *Usability heuristics for user interface design* von Nielsen angewendet werden [86]. Auch sollte bei der Entwicklung von Erklärungen auf *Shneiderman's Eight Golden Rules of Interface Design* geachtet werden [88].

Abschließend bieten die vorgestellten **10 Heuristiken für Erklärungsdesign** eine Liste von wichtigen Aspekten, die während der Entwicklung von Erklärungen berücksichtigt werden sollten. So kann auf Ergebnisse der Wissenschaft zur Auswirkung von Erklärungen auf ausgewählte externe Qualitätsaspekte aufgebaut werden und die insgesamt von *End Usern* wahrgenommene Softwarequalität erhöht werden.

Kapitel 6

Leitfadenanwendung

Dieses Kapitel behandelt die Anwendung des in Kapitel 5 vorgestellten Leitfadens zur Prüfung der Einsetzbarkeit in der Wirtschaft. Ziel war es bei Nutzung des Leitfadens zu überprüfen, ob dieser (i) verständlich ist und, (ii) ob mithilfe des Leitfadens Erklärungen in ein bestehendes System integriert werden können, die das Ziel der Integration erfüllen.

Das Kapitel ist in die Entwicklung von Erklärungen und die Evaluation der integrierten Erklärungen aufgeteilt. Der erste Teil enthält sowohl die Anforderungserhebung, den Designprozess und die technische Umsetzung. Im zweiten Teil werden die integrierten Erklärungen zunächst in einer Case-Study und später in einem Quasi-Experiment evaluiert.

6.1 Integration von Erklärungen anhand des Modells

Um eine praxisnahe Möglichkeit zu schaffen, wurde für die Anwendung des Leitfadens eine Kooperation mit der Firma *Graphmasters GmbH* eingegangen. *Graphmasters* ist ein in Hannover ansässiges Software-Unternehmen, welches vorwiegend Navigationssoftware entwickelt sowie Vekehrslenkungsmöglichkeiten und Tourenplanung für verschiedene Stakeholder (z. B. Logistikunternehmen und Verkehrsmanagementzentralen) anbietet.

Kernprodukt des Unternehmens ist eine Routing-Engine für den Straßenverkehr, welche auf Schwarmintelligenz basiert und damit alle Nutzer des Systems durch eine Cloud vernetzt. Der Algorithmus ermöglicht es, jedem Fahrzeug eine individuelle Route zuzuordnen und gibt damit jedem Teilnehmer im Schwarm die bestmögliche Route zu seinem Ziel. Diese Art von Routingalgorithmen wird „Kollaboratives Routing“ genannt. Aufgrund eines Reservierungssystems erhalten Fahrzeuge mit gleichen oder ähnlichen Zielen verschiedenen Routen. Dies führt dazu, dass die vorhandene Infrastruktur besser genutzt wird.

Der große Unterschied zu herkömmlichen Navigationssystemen („Egoistisches Routing“) ist, dass der von *Graphmasters* entwickelte Algorithmus das Entstehen von Staus durch die Verteilung der Autos auf verschiedene Strecken im Vorhinein verhindert. Bekannte Navigationslösungen reagieren nur auf eine bereits vorhandene Überfüllung von Straßen. Bei Erkennung einer Überfüllung würden jedoch alle Fahrzeuge mit dem Navigationssystem auf dieselbe Alternativroute gelenkt und der

Stau so nur verschoben werden. Die NUNAV genannte Technologie schafft es folglich durch Verteilungen, die Fahrzeit für alle Fahrzeuge zu verringern.

6.1.1 NUNAV Navigation

Der NUNAV-Routing-Algorithmus ist bei *Graphmasters* in verschiedene Produkte integriert. Eines der Produkte ist *NUNAV Navigation*. Dies ist eine frei verfügbare Navigationssoftware für Smartphones, deren Zielgruppe Endanwender sind, welche eine geführte Navigation nutzen wollen (mehr siehe Unterabschnitt 6.1.3). Mithilfe dieser können sich Privatanwender wie von bekannten Navigationslösungen gewohnt, zu beliebigen Zielorten navigieren lassen. Ferner haben Nutzer die Möglichkeit direkt nach Veranstaltungen und von NUNAV verwalteten Orten zu suchen. Dabei können Veranstalter NUNAV als individuelles Parkleitsystem einsetzen. Navigieren Nutzer mit *NUNAV Navigation* zu einem verwalteten Suchergebnis, werden sie auf einen freien Parkplatz geführt. Dabei kann auch zwischen verschiedenen Rollen unterschieden werden. So können Aussteller von Messen etwa direkt zum richtigen Eingang navigiert werden, statt auf einen Besucherparkplatz.

Für diese Arbeit dient die bestehende App *NUNAV Navigation* als Grundlage für die Integration von Erklärungen und ist somit das *System*, für das die Erklärbarkeit im Kontext der Anwendung des vorgestellten Leitfadens verbessert werden soll.

Technischer Überblick

Im Allgemeinen setzt die *Graphmasters GmbH* auf eine *Micro-Service*-Infrastruktur. Das bedeutet, die einzelnen Teilsysteme der NUNAV-Technologie sind stark gekapselt und werden unabhängig voneinander bei mehreren Cloud-Infrastruktur-Anbietern betrieben. So ist nicht nur eine unabhängige Entwicklung möglich, sondern auch die Skalierung einzelner Systeme ist einfach und kostengünstig umsetzbar. Clients für die Services sind entweder mobile Anwendungen für Smartphones oder Webanwendungen.

Abbildung 6.1 stellt abstrakt alle für diese Arbeit relevanten Services der Infrastruktur dar. Wichtig ist dabei zum einen der Routing-Algorithmus an sich (*Nugraph*), welcher die Daten für die Navigation bereitstellt. Zum anderen ist die mobile Anwendung von *NUNAV Navigation* relevant, da dort Erklärungen integriert werden sollen. Außerdem gibt es für jede Client-Anwendung einen sogenannten *Backend for Frontend* (BFF). Für die mobilen Anwendungen ist dies der *Mobile BFF*. Dieser stellt die einzige Kommunikationsschnittstelle der Clients mit allen anderen Services der Infrastruktur dar. In der Regel wird die Kommunikation zwischen den Services bzw. Clients und dem BFF über REST [1] -Schnittstellen mit JSON [2] als Übertragungsformat abgewickelt.

Nugraph *Nugraph* ist der Gesamtbegriff in der Architektur für die Services, welche zum Routing-Algorithmus gehören. Dieser berechnet auf der Basis eines prädiktiven Verkehrsmodells die individuell schnellste Route zwischen zwei Punkten.

[1] *Representational State Transfer* ist ein Paradigma zum Austausch von Daten über das HTTP-Protokoll.

[2] *JavaScript Object Notation* ist ein Datenformat, welches vor allem für Maschine-zu-Maschine-Kommunikation eingesetzt wird und Programmiersprachen unabhängig ist.

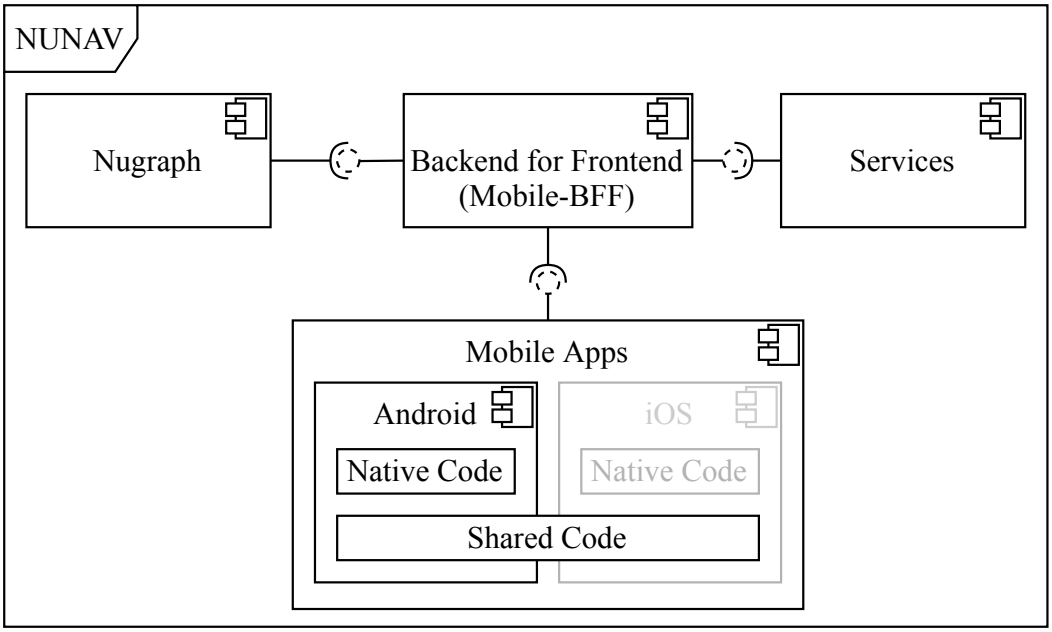


Abbildung 6.1: Ausschnitt aus der NUNAV Software Architektur (UML-Komponenten-Diagramm)

Dabei wird anhand von ca. 1,5 Millionen Verkehrsmessungen (z. B. *Floating Car Data*[3]) in der Minute der Verkehr auf der Route mithilfe von künstlicher Intelligenz vorausgesagt. Alle Daten, die zur Navigation (z. B. Verlauf und Geschwindigkeit auf der Route) benötigt werden, kommen von diesem Teilsystem.

Backend for Frontend Der in Abbildung 6.1 gezeigte *Mobile-BFF* ist zum einen für die korrekte Weiterleitung von Anfragen an den richtigen Service zuständig. Zudem bereitet dieser auch Daten anderer Services für die mobilen Apps auf oder konvertiert die verschiedenen Formate. Dabei werden zum Teil Daten mehrerer Services gebündelt oder neue Daten berechnet.

Mobile Apps Die mobilen Anwendungen für Smartphones bereiten zum einen die Daten der *Micro-Services* für die Nutzer auf und zeigen sie diesen an. Zum anderen verarbeiten diese lokal die vom BFF bereitgestellte Routen. Dabei berechnen die Applikationen die Position auf der Route, zum Geben von Abbiege-Kommandos und Anzeigen des Routenverlaufs. Die *Mobile Apps* werden für Android und iOS parallel in den jeweiligen vom Hersteller bereitgestellten Frameworks nativ entwickelt. Außerdem gibt es eine geteilte Code-Basis, welche Kotlin-Multiplatform[4] als Technologie nutzt.

Die drei vorgestellten Systemteile können entweder Daten für Erklärungen liefern oder sind für die Auslieferung der Erklärungen an sich zuständig. Im Rahmen

[3] *Floating Car Data* (FCD) sind mit Zeitstempeln versehene Geolokalisierungs- und Geschwindigkeitsdaten, die direkt von fahrenden Fahrzeugen erfasst werden.

[4] *Kotlin-Multiplatform* ist eine Technologie zur Kompilierung von Kotlin-Code in Binärdaten für verschiedene Plattformen und Prozessorarchitekturen.

dieser Arbeit wurden dabei sowohl auf dem BFF, als auch in der Mobilanwendung für Android Änderungen vorgenommen. Außerdem wurden Änderungen von der *Graphmasters GmbH* in *Nugraph* und einem weiteren Service vorgenommen, um benötigte Daten für die entwickelten Erklärungen zur Verfügung zu stellen.

6.1.2 Ermittlung des Erklärungsbedarfs

Die Anwendung des Leitfadens zur Integration von Erklärungen ist in mehrere Teile gegliedert. In einem ersten Schritt wurden die existierenden Verständnisprobleme in NUNAV analysiert. Diese wurden aus verschiedenen Datenquellen der *Graphmasters GmbH* zusammengefasst. Dazu zählen unter anderem mehrere Support-Kanäle.

Um diese Analyse anzureichern, wurde darauf aufbauend ein Workshop mit mehreren Mitarbeitern der *Graphmasters GmbH* durchgeführt. Ergebnis diesen Workshops waren dann eine Aufstellung der Probleme, Ziele und Rohanforderungen für Erklärungen und Ideen zur Umsetzung. Auf Basis dieser Ergebnisse wurden dann im Rahmen dieser Arbeit die Anforderungen konkretisiert und in Designs für Erklärungen umgesetzt. Abschließend ist die technische Realisierung in *NUNAV Navigation* erfolgt.

User-Feedback-Analyse

Für die initiale Analyse, welche Verständnisprobleme in *NUNAV Navigation* bestehen, wurde zunächst manuell das Feedback, welches im Google Play Store und Apple App Store von Nutzern in Form von App-Reviews gegeben wurde, analysiert (ab Oktober 2020). Daraus sind Themengebiete abgeleitet worden, und die Anzahl der Reviews, die ein Themengebiet ansprechen, wurden gezählt. Aus insgesamt 304 Reviews konnten 46 identifiziert werden, aus denen klare Themen abgeleitet werden können. 33 dieser Review thematisieren Verständnisprobleme. Die Themen, die häufig vorkamen und für Erklärbarkeit relevant sind, sind in Anhang A aufgelistet. Dabei wurde als Hauptproblem ein zu geringes Verständnis für den kollaborativen Routing-Algorithmus identifiziert. Allein fünf der 16 Reviews, aus denen sich dies ableiten lässt, bemängeln, dass NUNAV keine Alternativrouten vorschlägt. Dies weist darauf hin, dass nicht klar ist, dass das System des Verteilens der Nutzer auf die vorhandene Verkehrsinfrastruktur nur möglich ist, wenn die Nutzer den vorgeschlagenen Routen folgen und keine Alternativen nehmen.

Außerdem gibt es Hilfe-Artikel, die über einen *Hilfe-Center* der *Graphmasters*-Webseite erreichbar sind[5]. Dabei wurden die Klickzahlen mit den Review-Anzahlen verglichen, welche vom Verhältnis her ähnlich der Anzahlen an Reviews ist.

Aus den Reviews und Hilfe-Center-Artikeln wurden dann erste Themen beziehungsweise Nutzerfragen abgeleitet und mit dem Team „Solution Experts“ von *Graphmasters* diskutiert. Das Team ist für die Betreuung von Kunden und die Bearbeitung von Feedback zuständig. Die Ergebnisse sind in Tabelle 6.1 zu sehen.

Auf Basis dieser ersten Ergebnisse wurde im Anschluss ein Workshop mit mehreren *Graphmasters*-Mitarbeitern durchgeführt.

[5] https://support.graphmasters.net/, besucht: 01.10.21

Thema	Fragen
Kollaboratives Routing	Was ist / wie funktioniert kollaboratives Routing? Warum brauche ich eine ständige Internetverbindung? Wieso werden in NUNAV keine Standardrouten angezeigt?
Navigation	Warum wird diese Route / der Umweg genommen? Warum sind die Routen bei NUNAV länger? Warum ist die Position ungenau?
Vertrauen in das System	Woher kommen die Verkehrs-/ Sperrungsdaten? Was passiert mit meinen Daten? Wofür benötige ich Ortungsdienste beim App-Start? Kann die Güte der Route nicht bewerten?
Funktionen	Was zeigt die Rainbow-Route an? Wie kann ich eine Sperrung melden? Wie kann ich mein Ziel in NUNAV eingeben? Wie kann ich Favoriten anlegen?

Tabelle 6.1: Anzahl der Reviews im Google Play Store und Apple App Store mit mehrfach vorkommenden Themen

Workshop zur Integration von Erklärungen

Zum Festlegen der Ziele und Anforderungen an die Erklärungen sowie zum Sammeln erster Umsetzungsideen, wurde ein interdisziplinärer Workshop mit Frontend-Entwicklern, „Solution Experts“ und einem Produktmanager bei *Graphmasters* durchgeführt. Außerdem sollten Metriken gefunden werden, welche zur Überprüfung der Ziele verwendet werden können. Durch die direkte Anwendung des entwickelten Leitfadens im Workshop konnten außerdem Beobachtungen darüber, wie dieser eingesetzt wird, gesammelt werden. Dies war allerdings nicht Kern des Workshops.

Nach der initialen Vorstellung von Erklärbarkeit, dem Leitfaden für die Integration von Erklärungen sowie den Ergebnissen aus der User-Feedback-Analyse hat sich der Workshop an dem Aufbau des Modells für Erklärungen orientiert. Folglich wurden zuerst Informationen zu möglichen Nutzern (*Context*), dann Ziele (*Objectives*) und deren Umsetzungsmöglichkeiten (*Characteristics*) und schließlich die Evaluationsmöglichkeiten besprochen (*Evaluation*). Auch wenn der Workshop so geplant war, dass die einzelnen Aspekte nacheinander abgearbeitet werden, hat sich ergeben, dass die Ziele und deren Umsetzung gleichzeitig behandelt wurden. Grund dafür war, dass beim Aufstellen der Ziele und Rohanforderungen bereits die Umsetzbarkeit im Rahmen dieser Arbeit diskutiert wurde. Einige Fragen der Machbarkeit sind allerdings ungeklärt geblieben, da hierfür Wissen von den Entwicklern des Routing-Algorithmus benötigt wurde. Explizit ging es um die Frage, welche Daten für bestimmte Erklärungen überhaupt zur Verfügung gestellt werden können. Ein detailliertes Ergebnisprotokoll des Workshops ist in Abschnitt B zu finden.

Aus den Ergebnissen des Workshops wurden dann iterativ und in Rücksprache mit verschiedenen Teilnehmern des Workshops Personas, die konkreten Anforderungen sowie die finalen Erklärungen, die in *NUNAV Navigation* integriert werden sollten, entwickelt.

6.1.3 Personas von NUNAV-Nutzern

Um die Interpretation der Rohanforderungen aus dem durchgeführten Workshop besser interpretieren zu können, wurden Personas für typische NUNAV-Nutzergruppen abgeleitet. *Graphmasters* hat zwar kein genaues Bild darüber, welche Nutzertypen *NUNAV Navigation* nutzen, aus vergangenen Nutzergesprächen und Feedback haben sich allerdings zwei Gruppen herauskristallisiert, die klar benennbar sind.

Abbildung 6.2: Ayla - Lizensiert bei Shutterstock Nr. 1433736386

Erstnutzerin: Ayla ist 23 Jahre alt und studiert aktuell Mathematik im 7. Bachelorsemester. Sie wohnt in einem Studentenwohnheim und fährt jeden Tag mit der Bahn zur Universität. Sie kommt ursprünglich aus einem Dorf ca. 300 km entfernt und besucht ihre Eltern dort mindestens an drei Wochenenden im Monat. Dafür hat sie ein Auto, da sie mit ihrem Semesterticket nicht bis in die Heimat fahren kann. Sie wählt ihre Fahrtzeiten so, dass möglichst wenig Verkehr auf den Straßen ist.

In der Universität benutzt Ayla für Mitschriften und Notizen Zettel und Stift. Die Vorlesungsskripte druckt sie im Regelfall aus, um darin Notizen zu machen. Ihren Laptop nutzt sie nur, um die bereitgestellten Materialien der Dozenten herunterzuladen und mit ihnen zu kommunizieren. Folglich erledigt Ayla viele Aufgaben noch gerne analog statt digital.

Sie ist sehr aktiv auf Social-Media-Kanälen und postet regelmäßig Bilder mit ihrem Smartphone der neusten Generation.

Ayla ist großer Fan von Rap-Musik. Einer ihrer Lieblings-Rapper gibt in Kürze ein einzelnes Konzert in Deutschland und sie konnte zusammen mit zwei Freundinnen an Karten kommen. Da sie ein Auto hat, wollen die drei zusammen mit dem Auto zum Konzert fahren. Vom Veranstalter haben sie eine E-Mail erhalten, dass zur Anreise mit dem Auto möglichst die App *NUNAV Navigation* genutzt werden soll. Dies möchten sie ausprobieren.

Abbildung 6.3: Michael - Lizensiert bei Shutterstock Nr. 602783837

Vielfahrer: Michael, 34 Jahre alt, arbeitet als Unternehmensberater in einem großen Beratungsunternehmen. Er hat Betriebswirtschaftslehre studiert und ist seit dem Studium in der Firma.

Er wohnt mit seiner Frau und einer Tochter in der Vorstadt und muss jeden Morgen zur Hauptverkehrszeit in die Stadt fahren, um in sein Büro zu kommen. Dadurch steht er nahezu täglich auf dem Weg zur Arbeit im Stau. Die öffentlichen Verkehrsmittel kommen für ihn nicht infrage, da er mit ihnen mehr als doppelt so lange bräuchte.

Auch zu Vorortterminen beim Kunden fährt er ungefähr einmal pro Woche. Dabei interessiert ihn vor allem, wie lange er für die Strecke benötigt und dass er pünktlich ankommt.

Lange Zeit hat er das in sein Auto integrierte Navigationssystem verwendet, das aber nur vereinzelt auf Staus reagiert hat und ihn jeden Tag auf der gleichen Strecke zur Arbeit geschickt hat, wenn nicht ein besonderes Ereignis wie Unfälle oder Sperrungen auf der Strecke waren.

Michael hat ein modernes Dienst-Smartphone mit Android und ist generell Technik interessiert. Zwischendurch hat er die auf seinem Smartphone vorinstallierte Navigations-App verwendet, da diese den aktuellen Verkehr anzeigen kann. Diese leitete ihn zum Teil aus dem Stau des Berufsverkehrs raus, in dem er stand. Er hat aber meist ähnlich lange zur Arbeit gebraucht wie im Stau. Außerdem achtet Michael darauf, was mit seinen Daten passiert und da ist er sich bei dem Anbieter unsicher, wie anonym diese verarbeitet werden.

Seit einigen Tagen verwendet er die Navigations-App NUNAV Navigation, die verspricht ihn jeden Tag auf dem schnellsten Weg zur Arbeit zu bringen. Die App schlägt ihm ab und zu seiner Meinung nach „komische“ Routen vor. Anfangs hat er sich dann nicht an diese gehalten. Später hat er die NUNAV-Routen vollständig ausprobiert und war tatsächlich etwas schneller als sonst auf dem Weg zur Arbeit. Michael ist aber immer noch skeptisch, da er die vorgeschlagenen Routen nicht immer versteht und sich bei Umwegen daher immer noch auf seine Intuition verlässt.

6.1.4 Anforderungen der Erklärungen

Mithilfe der Rohanforderungen und Ziele aus dem Workshop sowie den erstellten Personas können sowohl funktionale als auch nicht-funktionale Anforderungen für Erklärungen in NUNAV entwickelt werden

Nicht-Funktionale Anforderungen

Als Methode zum Ableiten der konkreten NFRs wurde das Aufstellen eines Qualitätsmodells gewählt [32]. Das Modell ist in Abbildung 6.4 dargestellt. Bei der Erstellung wurde iterativ Rücksprache mit den verschiedenen Teilnehmern des Workshops gehalten. Als Basis für die Qualitätsziele wurden die in dem Modell für Erklärungen unter *Objectives* aufgeführten Ziele verwendet. Die Endfassung der Anforderungen wurde schlussendlich vom Teamleiter des „Mobile Application"-Teams abgesegnet.

Normalerweise findet eine systematisierte Anforderungserhebung, wie sie im Rahmen dieser Arbeit erfolgt ist, bei *Graphmasters* keine Anwendung. Die Rolle des *Requirements Engineers* hat sich daher als herausfordernd herausgestellt und ist in den agilen Prozessen von *Graphmasters* nicht vorgesehen. Insbesondere fehlte zum Teil das Verständnis, warum die Anforderungen in der folgend vorgestellten, konkreten Weise vorliegen sollen. In den Rücksprachen wurde deutlich, dass Ziele für die Produkte normalerweise deutlich abstrakter gehalten werden. Für eine Evaluation der Integration von Erklärungen im Rahmen dieser Arbeit werden die konkreten Anforderungen zur Prüfung allerdings benötigt.

In Abbildung 6.4 ist zu erkennen, dass für *Graphmasters* im Allgemeinen die Erhöhung der Nutzung und die Akzeptanz des Systems im Vordergrund stehen. Bei der Nutzung ist für *Graphmasters* die interessanteste Metrik die Anzahl der aktiven

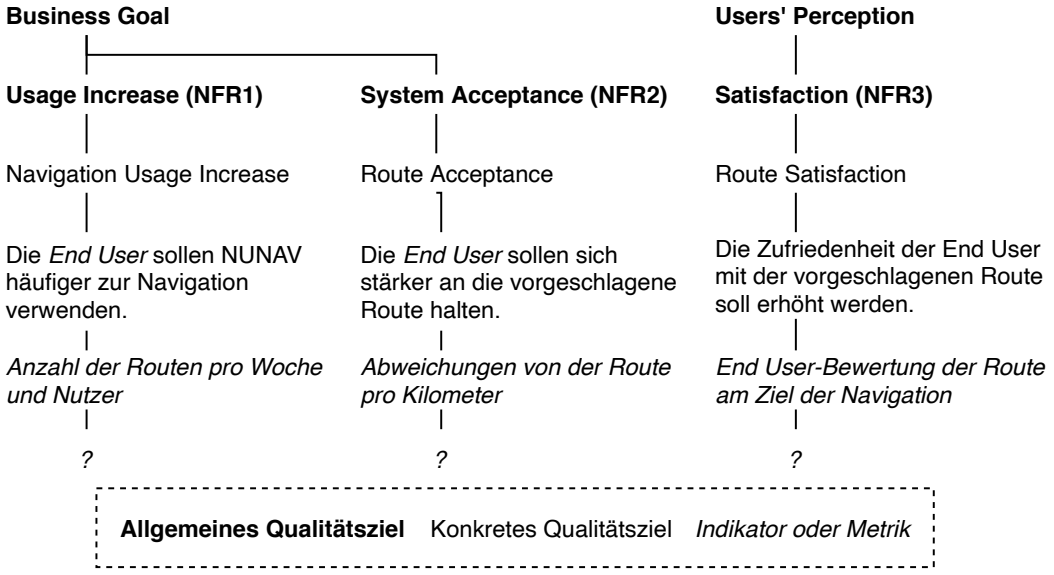


Abbildung 6.4: Qualitätsmodell für die Integration von Erklärungen in NUNAV Navigation

Nutzer. Da sich eine Messung von Veränderungen dieser durch die Integration von Erklärungen allerdings nicht über einen kurzen Zeitraum messen lässt, wurde als Metrik die Anzahl der Routen pro Nutzer und Woche vorgeschlagen und schließlich auch durch *Graphmasters* bestätigt. Für die Akzeptanz des Systems ist ein Ziel, dass sich die *End User* möglichst viel an die vorgeschlagene Route halten. Als Metrik wurde hierfür die Anzahl der Abweichungen von der Route pro Kilometer verwendet.

Ein weiteres wichtiges Ziel von *Graphmasters* ist die Zufriedenheit der *End User* mit den Routen. Hierfür bestand bereits vor dem Beginn dieser Arbeit eine Metrik. *End User* können bei Erreichen des Ziels einer Navigation auf einer Skala mit ein bis fünf Sternen angeben, wie zufrieden sie insgesamt mit der Route waren. Bisher wurde diese Metrik allerdings noch nicht systematisch ausgewertet.

Für alle Metriken fällt auf, dass in Abbildung 6.4 keine Sollwerte definiert sind. Aus den Gesprächen hat sich ergeben, dass für *Graphmasters* erstens nicht wichtig ist, um wie viel die verschiedenen Metriken sich verbessern, solange mit der Integration von Erklärungen eine Besserung eintritt. Außerdem ist sind keine Basiswerte vorhanden, auf die man sich beziehen kann. Daher wurden die konkreten Anforderungen in Bezug auf signifikante Unterschiede formuliert. Die Formulierungen orientieren sich dabei an gängigen Vorlagen, enthalten allerdings nicht die in den Vorlagen geforderten Sollwerte [89, 90]:

NFR1 Die Anzahl der gefahrenen Routen mit *NUNAV Navigation* pro Nutzer und Woche soll signifikant messbar erhöht werden.

NFR2 Die durchschnittliche Anzahl der Abweichungen der Nutzer pro Kilometer von der durch *NUNAV Navigation* vorgeschlagenen Route soll signifikant messbar verkleinert werden.

NFR3 Die durchschnittliche Bewertung der Route am Ende der Navigation in *NUNAV Navigation* soll signifikant messbar erhöht werden.

Funktionale Anforderungen

In diesem konkreten Fall stand die Integration von Erklärungen als Lösungsansatz zum Erreichen der Ziele bereits fest. Daher wurden die *Explanation Goals* aus dem Modell für Erklärungen ausgewählt, welche in Erklärungen umgesetzt werden sollten. Dabei steht vor allem auch die Frage im Raum, welche Aspekte genau erklärt werden müssen [19].

Ein Ergebnis des durchgeführten Workshops war die Verständnisprobleme der Nutzer, welche durch Erklärungen gelöst werden sollen und erste Umsetzungsideen. Aus diesen beiden Punkten konnten *Transparency* und *Scrutability* als Ziele für die Integration von Erklärungen abgeleitet werden.

Auch die Konkretisierung der funktionalen Anforderungen ist iterativ erfolgt. Eine Herausforderung dabei war vor allem, dass *Graphmasters* in der Regel keine konkreten Anforderungen formuliert, sondern direkt anhand von Ideen Mockups erstellt und dann iterativ mehrere Prototypen entwickelt.

Für die funktionalen Anforderungen wurde ein Zielbaum ähnlich zu Qualitätsmodellen aufgestellt, in dem die Konkretisierungsschritte zu sehen sind. Abbildung 6.5 stellt diesen dar.

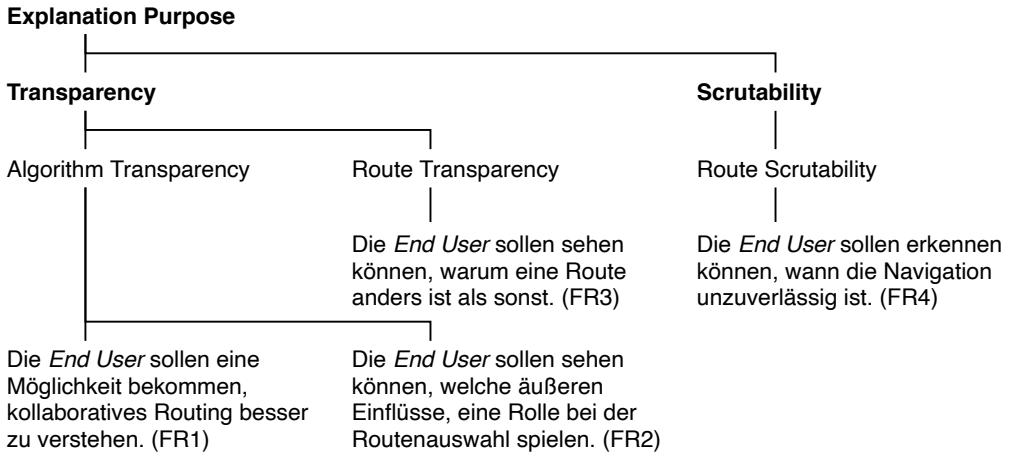


Abbildung 6.5: Konkretisierungsschritte bei der Entwicklung der funktionalen Anforderungen an Erklärungen in *NUNAV Navigation*

Aus den Konkretisierungsschritten (siehe Abbildung 6.5) wurden schlussendlich folgende funktionalen Anforderungen abgeleitet:

FR1 *NUNAV Navigation* muss den *End Usern* die Möglichkeit bieten, auf eine Erklärung zuzugreifen zu können, die den kollaborativen Routing-Algorithmus erklärt.

FR2 *NUNAV Navigation* muss den *End Usern* die Möglichkeit bieten, abzurufen, welche Informationen zu Verkehrsereignissen (z.B. Verkehrsfluss, Sperrungen, Baustellen, Staus) in den Algorithmus einfließen.

FR3 *NUNAV Navigation* muss den *End Usern* während der Navigation Informationen zum Verkehrsgeschehen auf der aktuellen Route liefern.

FR4 Wenn die Genauigkeit der Positionierung unzuverlässig ist, muss *NUNAV Navigation* den *End Usern* anzeigen, dass die Positionierung aktuell nicht zuverlässig ist.

Auf Basis dieser Anforderungen wurden dann Erklärungen für die Integration in *NUNAV Navigation* entwickelt.

6.1.5 Design der Erklärungen

Das Gestalten der Erklärungen ist mit mehreren Design-Mockups umgesetzt worden. Grundsätzlich ist für die funktionalen Anforderungen 1-4 jeweils ein Erklärungstyp entstanden. Wie aus Protokoll des durchgeführten Workshops zu entnehmen ist, wurden bereits in diesem einige Ideen für Erklärungen auf Basis der Vorschläge zur Umsetzung von Erklärungen aus dem Modell für Erklärungen entwickelt (*Characteristics*).

Außerdem wurden bei der Umsetzung die Zusammenhänge zwischen den Eigenschaften von Erklärungen auf Qualitätsaspekte sowie die Heuristiken für das Design von Erklärungen beachtet. Insbesondere wurde darauf geachtet, dass die

Erklärungen dezent sind, sodass sie möglichst keinen negativen Einfluss auf die *Usability* haben. Längere Erklärungen werden nur optional eingesetzt. Auch wurde, wie im Leitfaden vorgeschlagen, versucht, für eine Information möglichst kombinierte Stile zu nutzen, sodass die *End User* diese auf mehrere Weisen erfassen können.

Da eine weitere Empfehlung des Leitfadens die *Context Sensitivity* von Erklärungen betrifft, wird im Folgenden der *Context* von *NUNAV Navigation* beschrieben. Grundsätzlich kann im Fall von Navigationsanwendungen zwischen zwei verschiedenen Situationen unterschieden werden. Die Erste ist die Nutzung der Anwendung außerhalb der aktiven Navigation. Dies beinhaltet das Auswählen des Ziels, eine Routenübersicht sowie die Möglichkeit des Startens der Navigation. Hier haben die *End User* keinen Zeitdruck und können sich auf die Interaktion mit der Applikation konzentrieren. Trotz dessen wurden die Erklärungen so entwickelt, dass diese von den *End Usern* wie von Chazette, Karras und Schneider sowie Wang vorgeschlagen, optional angefordert werden müssen [7, 21]. Folglich wurden für die Anforderungen FR1 und FR2 in diesem *Context* umgesetzt. Grund dafür ist, dass die Informationsmenge zur Erklärung des Routing-Algorithmus und der dazugehörigen Daten so groß ist, dass eine Integration in der zweiten möglichen Situation nicht ohne Beeinflussung der Nutzerperformanz möglich ist.

Die zweite Situation ist die aktive Navigation. Während dieser sollten die *End User* so wenig wie möglich abgelenkt werden, da sie der vollen Konzentration auf die Straße bedürfen. Daher sind alle Erklärungen, die während der Navigation den *End Usern* zur Verfügung gestellt werden, entgegen der Empfehlung für Interaktionen mit Erklärungen ohne eine Manipulationsmöglichkeit durch die *End User* gestaltet. Da die Anforderungen FR3 und FR4 von einer aktiven Navigation abhängig sind, wurden diese im zweiten *Context* integriert.

Die vier entwickelten Erklärungstypen werden im Folgenden im Detail vorgestellt.

Kollaboratives Routing

[FR1] *NUNAV Navigation* muss den *End Usern* die Möglichkeit bieten, auf eine Erklärung zuzugreifen zu können, die den kollaborativen Routing-Algorithmus erklärt.

Wie bereits erläutert, ist der Kerngedanke des *NUNAV*-Routing-Algorithmus, dass jeder *End User* die für ihn individuell schnellste Route erhält und dabei nicht zwangsweise Hauptverkehrsstraßen nutzt. Folglich ist für diese erste Anforderung eine Erklärung entworfen worden, die das kollaborative Routing erklärt.

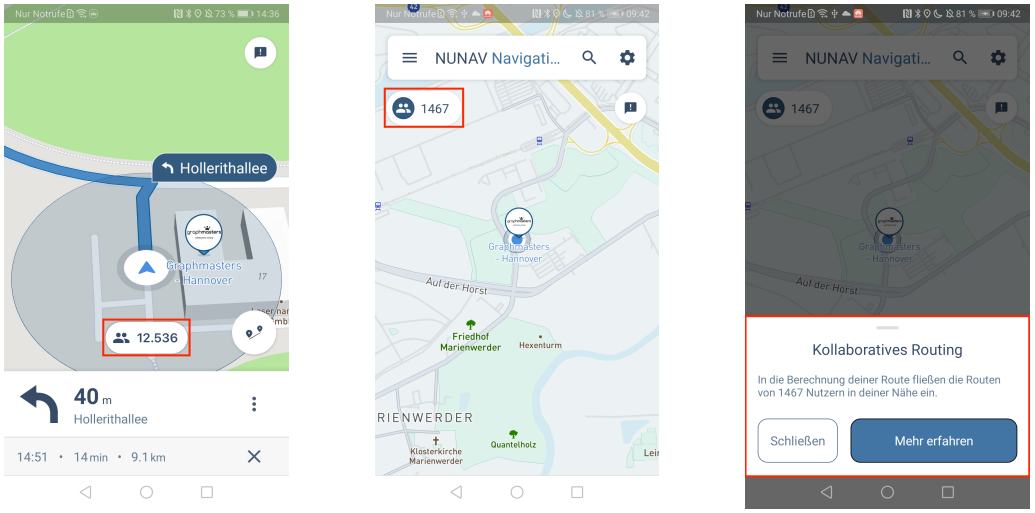
Wie bei genauerem Einblick in das Feedback von *End Usern* auffällt, ist eines der Grundprobleme, dass ihnen das Verständnis fehlt, dass sie vernetzt mit anderen Nutzern von *NUNAV* kollaborativ navigiert werden. Da dies einer der Hauptunterschiede zu anderen Navigationsanbietern ist, sind Erklärungen, die diesen Punkt betreffen, hauptsächlich für Erstnutzer wie Ayla wichtig. Um den Nutzern einen Einblick in diese Abhängigkeit von anderen Verkehrsteilnehmern zu geben, wird die aktuelle Anzahl der Nutzer, welche die eigene Route beeinflussen, angezeigt. Die Berechnung dieser Anzahl ist eine Annäherung, da sich nicht genau bestimmen lässt, welche Fahrzeuge auf der Straße einen realen Einfluss auf die eigene Navigation haben. Die Annäherung wurde allerdings als ausreichend befunden, da

die Richtigkeit der Erklärung, wie im Leitfaden beschrieben wird, keinen Einfluss auf die *Transparency* hat. In einem ersten Entwurf wurden noch alle aktiven Nutzer im System als Zahl angenommen. Dies wurde allerdings verworfen, da Nutzer aus dieser Zahl wenig Informationen über den unmittelbaren Einfluss auf für eigene Navigation erhalten.

Die Anzahl der Nutzer wurde von einem Backend-Team bei *Graphmasters* zur Verfügung gestellt, die restliche Umsetzung in *BFF* und App ist im Rahmen dieser Arbeit erfolgt. Für die Anzeige der Nutzerzahl wurden verschiedene Positionen im User Interface ausprobiert, wie in Abbildung 6.6 zu erkennen ist. Für die finale Version ist die Entscheidung gefallen, da an dieser Stelle bereits andere Informationen als sogenanntes *Growl* angezeigt werden und diese Stelle zur Anzeige kurzer Informationen somit im Rahmen der *Usabiltiy* bereits erfolgreich ohne negatives Feedback genutzt wird. Die integrierten Erklärungen sind jeweils rot umrahmt.

Die Erklärung des Routing-Algorithmus ist in zwei verschiedenen Granularitätsstufen umgesetzt worden. Tippen *End User* auf das *Growl* mit der Anzahl der Nutzer, gelangen sie zu einer kurzen Erklärung (siehe Abbildung 6.6, (c)). Ursprünglich war der Text „In den letzten 15 Minuten haben wir anonyme Verkehrsdaten von *<Anzahl der Nutzer>* Nutzern in deiner Umgebung erhalten.“. Dieser Satz wurde in internem Feedback bei *Graphmasters* allerdings für zu technisch bzw. datengetrieben befunden. Daher wurde er so umformuliert, dass der direkte Einfluss auf die *End User* klar wird (siehe Abbildung 6.6, (c)).

Als weitere Erklärungsmöglichkeit können die *End User* mittels „Mehr erfahren“ zu dem entsprechenden Hilfe-Center-Artikel springen (siehe Abschnitt B). Eine Verknüpfung aus der App heraus direkt zum Hilfe-Center gab es vor dieser Arbeit in *NUNAV Navigation* noch nicht.



(a) 1. Prototyp zur Positionierung der Nutzerzahl

(b) Finales Design der Positionierung der Nutzerzahl

(c) Finales Design der kurzen Erklärung zum kollaborativen Routing

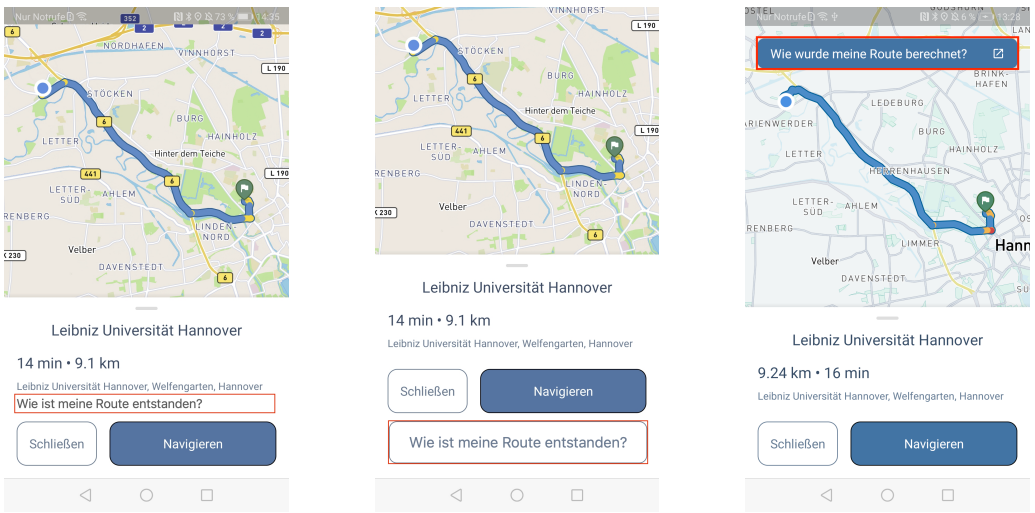
Abbildung 6.6: Prototyp und finale Designs für die Erklärung zum kollaborativem Routing

Einflüsse auf die Routenberechnung

[FR2] *NUNAV Navigation* muss den *End Usern* die Möglichkeit bieten, abzurufen, welche Informationen zu Verkehrsereignissen (z.B. Verkehrsfluss, Sperrungen, Baustellen, Staus) in den Algorithmus einfließen.

Auf der Karte, welche das Hauptinteraktionselement von *NUNAV Navigation* darstellt, sind Sperrungen, Baustellen und Staus bereits eingezeichnet. Wie aus dem Nutzerfeedback hervorgeht, reichen diese *Context*-Informationen für das Verständnis der *End User* nicht aus. Daher wurde ein neuer Hilfe-Center-Artikel im Rahmen dieser Arbeit zu diesem Thema angelegt (siehe Abschnitt B). Dieser Artikel ist für Nutzer geeignet, welche NUNAV zum ersten Mal wie Ayla oder noch nicht lange wie Michael verwenden. So können sie sich anhand der Daten auf der Karte im Anschluss an das Lesen des Artikels die verschiedenen Routen verständlicher erklären.

Dieser Beitrag wurde ebenfalls vor dem Start der Navigation erreichbar gemacht. Um die Standardkartenansicht nicht durch mehrere Möglichkeiten zum Aufrufen von Hilfe-Center-Artikeln zu überladen, wurde die Möglichkeit, zu dieser Erklärung zu gelangen, in die Routen-Vorschau integriert. Auch für diese wurden mehrere Prototypen erstellt, wobei bei der finalen Entscheidung darauf geachtet wurde, dass die Aufrufmöglichkeit nicht zu viel Platz einnimmt und der Text neugierig macht. Die verschiedenen Prototypen sind in Abbildung 6.7 zu sehen.



(a) 1. Prototyp zur Positionierung des Aufrufs der Erklärung

(b) Alternativer Prototyp zur Positionierung des Aufrufs der Erklärung

(c) Finales Design zur Positionierung des Aufrufs der Erklärung

Abbildung 6.7: Prototyp und finale Designs für den Aufruf der Erklärung zu Einflüssen auf die Routenberechnung

Verkehrsaufkommen

[FR3] *NUNAV Navigation* muss den *End Usern* während der Navigation Informationen zum Verkehrsgeschehen auf der aktuellen Route liefern.

Wie aus der Persona von Michael hervorgeht, wundert er sich zum Teil, warum *NUNAV Navigation* aus seiner Sicht „komische“ Routen vorschlägt. Ein Erklärungsansatz ist, das Verkehrsgeschehen auf der aktuellen Route anzuzeigen. Diese *Context*-Information soll dabei helfen zu verstehen, warum *NUNAV Navigation* andere Routen wählt.

Dies ist allerdings nicht trivial, da die Berechnung des Verkehrsaufkommens und dessen subjektive Wahrnehmung der *End User* nicht klar ist. Ein weiteres Problem ist, dass es nicht möglich ist, zu berechnen, welche Route für Nutzer auf der gleichen Strecke subjektiv als „normal“ empfunden wird. Nachdem mehrere Lösungsansätze für dieses Problem in kleinen Runden bei *Graphmasters* diskutiert wurden, wird als Lösung die Berechnung des Verhältnisses von durchschnittlicher Dauer der Route zu aktueller Dauer als Basiswert genommen. Die Daten werden dabei durch *Nugraph* zur Verfügung gestellt und im Rahmen dieser Arbeit im *BFF* auf die Stufen „wenig Verkehr“, „mäßiger Verkehr“ und „viel Verkehr“ abgebildet. Außerdem gibt es eine „normale“ Verkehrssituation, in der den *End Usern* keine Erklärung angezeigt wird. Die Abbildungsfunktion ist in den Zusatzmaterialien zu finden. Die Schwellwerte wurden intern durch Testfahrten ermittelt.

Die Informationen können *End User* auf drei Wegen erhalten. Zunächst werden die Informationen in der Routenvorschau angezeigt. Dabei wird die Gesamtfahrzeit für die Route je nach Verkehrsaufkommen grün (wenig Verkehr), Standardtextfarbe (normaler Verkehr), orange (mäßiger Verkehr) und rot (viel Verkehr) dargestellt. Da im ersten Prototypen aufgefallen ist, dass die Farben alleine nicht aussagekräftig sind (siehe Abbildung 6.8, (a)), wurde zusätzlich ein kurzer Erklärungstext eingefügt (siehe Abbildung 6.8, (b)). Dieser ist bei „normalem“ nicht zu sehen. So lernen die *End User* mit der Zeit die Bedeutung der Farben. Eine Übersicht mit allen Prototypen ist in Abschnitt B zu finden.

Außerdem werden während der gesamten Navigation die Ankunftszeit und die Fahrzeit in der entsprechenden Farbe angezeigt und bei sich ändernder Verkehrssituation auch aktualisiert.

Des Weiteren ist das Sprachkommando, welches bei Start der Navigation gegeben wird, je nach Verkehrssituation unterschiedlich (siehe Tabelle 6.2).

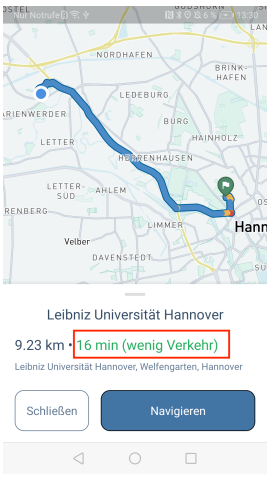
GPS-Qualität

[FR4] Wenn die Genauigkeit der Positionierung unzuverlässig ist, muss *NUNAV Navigation* den *End Usern* anzeigen, dass die Positionierung aktuell nicht zuverlässig ist.

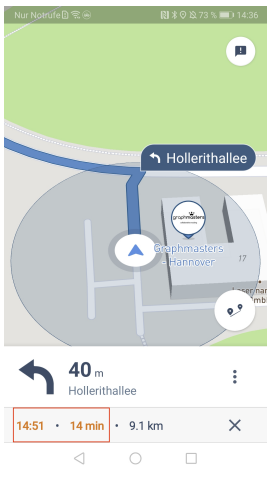
Die letzte entwickelte Erklärung betrifft den Aspekt der *Scrutability*. Ziel ist es *End Usern* anzuzeigen, wenn sie sich aktuell auf die Position auf der Route nicht verlassen können. Dies ist zum Beispiel in Tunneln der Fall und besonders kritisch, wenn in Kürze ein Abbiege-Kommando erfolgt und somit an der falschen Stelle kommt. Bewusst wurde auf die technische Bezeichnung „GPS“ verzichtet.

Für die Entscheidung, wann das GPS ungenau ist, wurde innerhalb der *NUNAV Navigation*-App ein neuer Algorithmus integriert, der diese trifft. Dabei spielt nicht

(a) Prototyp zur Darstellung des aktuellen Verkehrsaufkommens in der Routenvorschau



(b) Finales Design zur Darstellung des aktuellen Verkehrsaufkommens in der Routenvorschau



(c) Finales Design zur Darstellung des aktuellen Verkehrsaufkommens während der Navigation

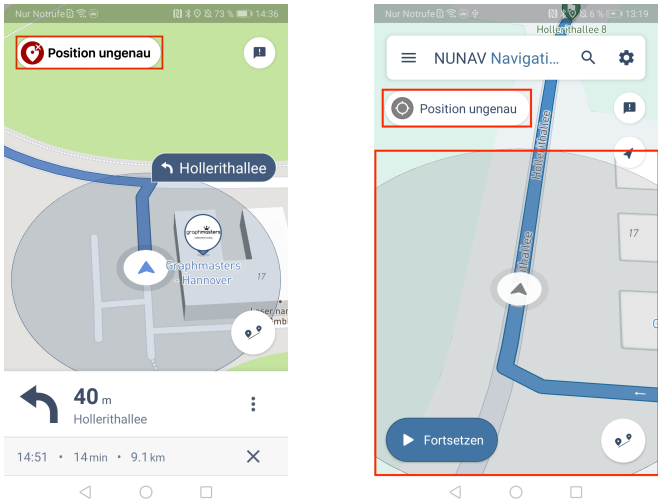
Abbildung 6.8: Prototyp und finale Designs für die Erklärung zum kollaborativem Routing

Verkehrsaufkommen	Sprachkommando
wenig	Heute sind die Straßen frei. Du wirst dein Ziel <Zielname> um <Uhrzeit> erreichen.
normal	Du wirst dein Ziel <Zielname> um <Uhrzeit> erreichen.
mäßig	Da heute etwas mehr los ist, wirst du dein Ziel <Zielname> um <Uhrzeit> erreichen.
viel	Da heute sehr viel los ist, wirst du dein Ziel <Zielname> um <Uhrzeit> erreichen.

Tabelle 6.2: Sprachkommandos zum Start der Navigation abhängig von der Verkehrssituation

nur die Genauigkeit, welche durch das GPS-Modul des Gerätes ermittelt wird, eine Rolle, sondern auch die Zeit, wann das letzte GPS-Update kam, da bei sehr schlechtem GPS keine Positionsupdates mehr vom Gerät geliefert werden. Der Algorithmus ist in den Zusatzmaterialien zu finden.

Als Design wurde im ersten Entwurf lediglich ein *Growl* angezeigt und die Sprachansage „Position ungenau“ ausgegeben. Insbesondere, wenn zum gleichen Zeitpunkt weitere *Growls* angezeigt werden und die Sprachansagen vom *End User* generell deaktiviert sind, hat sich die Sichtbarkeit bei Testfahrten als ungenügend herausgestellt. Daher wurde im finalen Entwurf zusätzlich das Positions-Icon auf der Karte und der Kreis, welcher die Genauigkeit der Position anzeigt, ausgegraut (siehe Abbildung 6.9, (b)).



(a) Prototyp zum Design bei schlechtem GPS

(b) Finales Design bei schlechtem GPS

Abbildung 6.9: Prototyp und finales Design für die Anzeige von schlechtem GPS während der Navigation

6.1.6 Umsetzung

Alle zuvor erläuterten und im Rahmen der vorliegenden Arbeit entwickelten Erklärungen wurden entweder im *BFF* oder innerhalb der *NUNAV Navigation* Android App umgesetzt. Die Integration ist nur auf Android erfolgt, da die Anzahl der Nutzer auf Android etwa 92% ausmacht und die Integration in die iOS-App den Aufwand auf Frontend-Seite verdoppelt hätte. Außerdem sind die Apps insgesamt auf verschiedenen Design-Ständen und können daher in einer Evaluation nicht direkt verglichen werden. Die Erklärungen wurden innerhalb des im Mobile-Team bei *Graphmasters* üblichen Entwicklungsprozesses in die App integriert. Grundsätzlich basiert dieser auf Scrum mit einwöchigen Sprints und *Github-Issues* als Tool zum Tracken von Aufgaben. Für jede abgeschlossene *Issue* erfolgt darüber hinaus ein Review durch ein anderes Team-Mitglied.

Die beschriebenen Erklärungen sind in allen Live-Versionen der Apps potenziell verfügbar. Allerdings wurde in dieser Arbeit ein *Feature-Flag*-System entwickelt, mithilfe dessen es möglich ist über den *BFF* einzelne Features für bestimmte *End User* zu aktivieren oder deaktivieren. Daher sehen nicht alle *End User* die Erklärungen.

Für die Integration der Erklärungen wurden Code-Änderungen am *BFF*, dem *Shared Code* der Mobilplattformen, sowie am nativen Code der Android-App im Rahmen dieser Arbeit vorgenommen. Der Großteil des Codes ist in den Zusatzmaterialien dieser Arbeit einsehbar. Wie zuvor erwähnt wurden zum Teil von anderen Teams bei *Graphmasters* zusätzliche Daten für die Erklärungen zur Verfügung gestellt. Diese Änderungen sind in dieser Arbeit nicht enthalten.

6.2 Evaluation der integrierten Erklärungen

Um die vorgestellten Erklärungen für *NUNAV Navigation* auf die Erfüllung der aufgestellten Ziele zu überprüfen, wurde anhand des entwickelten Leitfadens eine Studie zur Evaluation der Erklärungen entworfen. Zusätzlich werden die Ergebnisse aus dem durchgeführten Workshop in das Studiendesign mit einbezogen (siehe Abschnitt B). Die Studie ist in zwei Teile gegliedert. Im ersten Teil ist eine *Case Study* [34] durchgeführt worden, welche in die Produktivversion von *NUNAV Navigation* integriert wurde. Um die Ergebnisse daraus besser einordnen zu können, ist im Anschluss daran mit vier Nutzern ein Quasi-Experiment [34] durchgeführt worden. Die Teilnehmer wurden unabhängig von der vorherigen Studie ausgewählt.

6.2.1 Ziel der Evaluation

Das Ziel der Evaluation ist wie auch das Ziel der gesamten Arbeit (siehe Abschnitt 3.1) auf Basis der Vorlage von Wohlin et al. formuliert [34].

> Die Studie **analysiert** die integrierten Erklärungen **in Bezug auf** die externe Qualität des Systems **zur** Evaluation **aus der Sicht** von *End Usern* **im Kontext** der Benutzung von *NUNAV Navigation*.

Dem zu Folge sollte im Rahmen der Evaluation überprüft werden, ob die zuvor aus den abgeleiteten Zielen von *Graphmasters* definierten Qualitätsanforderungen durch die entwickelten Erklärungen erfüllt werden können.

Hypothesen

Die Überprüfung der Erfüllung der gestellten Qualitätsanforderungen erfolgt im Rahmen von Hypothesentests [34]. Grundsätzlich wurden zwei abstrakte Einflusshypothesen für die Integration der Erklärungen aufgestellt:

- WENN die Studienteilnehmer Erklärungen erhalten, DANN ist ein positiver Einfluss auf die externe Softwarequalität messbar im Vergleich zu Teilnehmern, welche keine Erklärungen erhalten.
- WENN die Studienteilnehmer mehrere Erklärungen erhalten, DANN ist ein positiver Einfluss auf die externe Softwarequalität messbar im Vergleich zu Teilnehmern, welche nur eine Erklärung erhalten.

6.2.2 Studienaufbau

Ein Ergebnis des durchgeführten Workshops war, dass die Erklärungen nach Möglichkeit unter Realbedingungen evaluiert werden sollen. Daraus ist die Anforderung abgeleitet worden, dass eine Studie die *End User* von *NUNAV Navigation* so wenig wie möglich bemerken sollten. Daher kommen lediglich Verhaltensmetriken oder bereits in der Anwendung integrierte Metriken zum Einsatz.

Die genutzten Metriken wurden aus dem Unterabschnitt 6.1.4 zur Definition der Anforderungen übernommen und sind bereits vor Beginn dieser Arbeit in *NUNAV Navigation* integriert gewesen.

Aufgrund von Einflusshypothesen durch Störgrößen wurden neben den Erklärungsmetriken zusätzliche Metriken in die Anwendung integriert. Das Ziel dabei war es, auf der Basis möglicher Einflüsse Datensätze herausfiltern zu können, welche die Ergebnisse der Analyse der Einflüsse durch integrierte Erklärungen zu verzerren (siehe Abschnitt 6.2.4).

6.2.3 Studiendurchführung

Insgesamt wurde folglich eine zweiwöchige Case Study als empirische Strategie gewählt. Um die Unterschiede in Bezug auf die externe Qualität messen zu können, wurden nicht allen Teilnehmern die Erklärungen angezeigt. Zusätzlich sollte nicht nur eine Veränderung durch die Erklärungen messbar gemacht werden, sondern auch zwischen den einzelnen entwickelten Erklärungen unterschieden werden. Trotz dessen sollte auch die Kombination der verschiedenen Erklärungen überprüft werden. Um jedoch für jede Bedingung genug Nutzer als Datengrundlage zu haben, wurden nicht alle verschiedenen Kombinationsmöglichkeiten der Erklärungen überprüft.

Grundsätzlich wurden die Erklärungen in die zwei Gruppen *statische* und *Context*-abhängige Erklärungen gegliedert. Unter den statischen Erklärungen sind die beiden Erklärungen, welche den Routing-Algorithmus und die Einflüsse auf die Routenberechnung erklären, zusammengefasst (siehe Unterabschnitt 6.1.5).

Analog dazu sind die *Context*-Erklärungen zum aktuellen Verkehrsaufkommen und zu Positionsungenauigkeiten während der Navigation als eine Bedingung für die Studie zusammengefasst worden (siehe Abschnitt 6.1.5 und Abschnitt 6.1.5).

Insgesamt sind folglich vier verschiedene Studiengruppen entstanden:

- **Gruppe 1**: Nutzer, die keine Erklärungen erhalten
- **Gruppe 2**: Nutzer, die statische Erklärungen erhalten
- **Gruppe 3**: Nutzer, die *Context*-abhängige Erklärungen erhalten
- **Gruppe 4**: Nutzer, die statische als auch *Context*-abhängige Erklärungen erhalten

Für die Zuordnung der einzelnen Studiengruppen wurde das im Rahmen dieser Arbeit entwickelte *Feature-Flag*-System verwendet. Anhand von zufällig generierten Identifikatoren wurde zu diesem Zweck ein Hashwert berechnet und die Teilnehmer anhand der Teilbarkeit durch vier, den einzelnen Gruppen zugeordnet (siehe Zusatzmaterialien).

6.2.4 Studienergebnisse

Im Folgenden werden die Ergebnisse der beschriebenen Case-Study vorgestellt. Nach einem Überblick über die Studienteilnehmer werden die Einflusshypothesen durch Störgrößen außerhalb von Erklärungen analysiert. Darauffolgend werden die abstrakten Hypothesen (siehe Abschnitt 6.2.1) jeweils auf die in Unterabschnitt 6.1.4 vorgestellten Anforderungen übertragen und im Anschluss geprüft. Für die Überprüfung der Hypothesen wird jeweils die Wahrscheinlichkeit p berechnet, dass ein Unterschied der Daten nur zufällig zustande gekommen ist. Für alle Wahrscheinlichkeiten wird für $p < 0.05$ angenommen, dass ein signifikanter Unterschied zwischen zwei Bedingungen vorliegt [vgl. 34].

Übersicht

Im Zeitraum der Studie haben insgesamt 9 745 *End User* Daten, welche die benötigten Metriken enthalten, beigetragen. Dabei sind diese 41 540 Routen gefahren. Dies enthält allerdings auch Daten, von *End Usern*, die keine aktive Navigation durchgeführt haben. Um diese Daten herauszufiltern, wurden Routen, bei denen weniger als 5 Kilometer zurückgelegt wurden, nicht berücksichtigt. Schlussendlich sind im zur Analyse verwendeten Datensatz folglich 4 012 *End User* mit 16 531 gefahrenen Routen enthalten. Für 625 der Routen liegt darüber hinaus eine Bewertung durch *End User* vor (siehe Tabelle 6.3).

Die Studienteilnehmer der Gruppen 3 und 4 erhalten während der Navigation so lange keine Erklärung, wie auf der aktuellen Route keine besonderen Vorkommnisse auftreten. Dies ist der Fall, wenn das Verkehrsaufkommen „normal“ ist und die *End User* während der gesamten Fahrt guten GPS-Empfang haben (siehe Abschnitt 6.1.5 und Abschnitt 6.1.5). Bei der Betrachtung von einzelnen Routen werden die Teilnehmer folglich in die Gruppen 1 oder 2 umsortiert, falls sie während der Navigation keine Erklärung erhalten haben. Bei der Betrachtung von Metriken über mehrere Routen werden Teilnehmer den entsprechenden Gruppen zugeordnet, wenn NUNAV ihnen mindestens einmal eine Erklärung aus der Studiengruppe angezeigt hat. Eine Übersicht findet sich in Tabelle 6.3.

Einflüsse außerhalb von Erklärungen

Zunächst wird die Hypothese geprüft, dass eine schlecht vorausgesagte Ankunftszeit (*Estimated Time of Arrival (ETA)*) im Vergleich zur wirklichen Ankunftszeit (*Actual*

Studiengruppe	Anzahl der Nutzer	Anzahl der Routen	
		Insgesamt	Mit Nutzerbewertung
Gruppe 1	1 778	4 807	133
Gruppe 2	1 397	3 413	135
Gruppe 3	468	4 571	184
Gruppe 4	369	3 740	173
Insgesamt	4 012	16 531	625

Tabelle 6.3: Übersicht über die Daten der Studiengruppen

Time of Arrival (ATA)) einen negativen Effekt auf die Anzahl der Abweichungen von der Route sowie die Zufriedenheit mit der Route hat. Die gleichen Auswirkungen werden für ein häufiges Auftreten von Positionsungenauigkeiten bei der Navigation untersucht. Dies wird analysiert, da ein negativer Zusammenhang vermutet wird, der die Ergebnisse der Untersuchung beeinflussen könnte. Überprüft werden die folgenden Hypothesen innerhalb der Kontrollgruppe (Gruppe 1: Ohne Erklärungen):

H1.1 WENN die *ATA* mehr als 10 % und mindestens 2 Minuten von der *ETA* abweicht, DANN hat dies einen signifikant messbaren negativen Einfluss auf die Zufriedenheit der Nutzer mit der Route.

H1.2 WENN die *ATA* mehr als 10 % und mindestens 2 Minuten von der *ETA* abweicht, DANN ist die Anzahl der Routenabweichungen signifikant messbar höher.

H1.3 WENN *NUNAV Navigation* auf der betrachteten Route pro 5 km durchschnittlich mindestens eine Positionsungenauigkeit aufwies, DANN hat dies einen signifikant messbaren negativen Einfluss auf die Zufriedenheit der Nutzer mit der Route.

H1.4 WENN *NUNAV Navigation* auf der betrachteten Route pro 5 km durchschnittlich mindestens eine Positionsungenauigkeit aufwies, DANN ist die Anzahl der Routen-Abweichungen signifikant messbar höher.

Die statistische Prüfung der Auswirkung von schlecht vorausgesagter Ankunftszeit auf die Anzahl der Routen-Abweichungen erfolgt mittels eines Kruskal-Wallis-Tests. Dieser wird verwendet, da die Datensätze für die beiden Gruppen verschieden lang und nicht normalverteilt (geprüft mit Shapiro-Wilk-Test) und zudem unabhängig voneinander sind. Das Ergebnis zeigt keinen Haupteffekt ($p = 0.197648$). Gleiches gilt für die Überprüfung eines Effektes auf die Nutzerzufriedenheit ($p = 0.564911$). Folglich können die Hypothesen H1.1 und H1.2 abgelehnt werden.

Für die Überprüfung, ob eine schlechte Positionierung einen Effekt auf die Nutzerzufriedenheit hat, ergibt sich anhand eines Kruskal-Wallis-Tests, dass kein signifikanter Effekt vorliegt ($p = 0.269231$). Folglich muss Hypothese H1.3 abgelehnt werden. Bei der Prüfung von Hypothese H1.4 kann festgestellt werden, dass die Anzahl der Routen-Abweichungen signifikant höher ist, wenn im Durchschnitt mehr als ein mal pro 5 km eine Positionsungenauigkeit während der Navigation auftritt ($p = 3.426601e - 15$). Folglich kann Hypothese H1.4 angenommen werden. Da eine häufige ungenaue Positionierung also bereits einen Einfluss auf die Anzahl der Abweichungen von der Route hat, werden Daten mit ungenauer Positionierung bei der Auswertung vom Einfluss von Erklärungen auf die Anzahl der Routen-Abweichungen herausgefiltert.

Häufigkeit der Nutzung

Bei der Analyse der Nutzungshäufigkeit von NUNAV wurde über den zweiwöchigen Studienzeitraum gemessen, wie viele Routen pro Nutzer und Studiengruppe in diesem Zeitraum gefahren wurden. Für die Überprüfung, ob die Erklärungen die Qualitätsanforderung für eine Erhöhung der Nutzung (NFR1) erfüllen, werden folgende Hypothesen aufgestellt:

H2.1 WENN Teilnehmer Erklärungen erhalten, DANN verwenden sie *NUNAV Navigation* signifikant häufiger, als wenn sie keine erhalten.

H2.2 WENN Teilnehmer nur einen der beiden vorgestellten Erklärungstypen erhalten, DANN nutzen sie *NUNAV Navigation* signifikant seltener, als wenn sie beide Erklärungstypen erhalten.

Abbildung 6.10 enthält eine Übersicht der Verteilung der Anzahl der gefahrenen Routen pro Nutzer für jede Studiengruppe. Für die weitere Analyse werden zunächst Ausreißer herausgefiltert. Dabei werden alle Datenpunkte, welche mehr als dreimal die Standardabweichung vom Median entfernt sind, aus dem Datensatz genommen [91]. Final sind 3 937 Nutzer im Datensatz für die Prüfung der Hypothesen verblieben. Daraus sind die in Tabelle 6.4 aufgelisteten Ergebnisse berechnet worden.

Zur Prüfung der Signifikanz wird wieder der Kruskal-Wallis-Test eingesetzt, da auch die Anzahl der gefahrenen Routen pro Studienteilnehmer nicht normalverteilt ist (Shapiro-Wilk: $p = 0.0$). Mit $p = 1.311644e-15$ als Ergebnis des Kruskal-Wallis-Tests kann davon ausgegangen werden, dass ein Haupteffekt vorliegt.

Studiengruppe	Mittelwert [N]	Standardabweichung [N]
Gruppe 1	3.2234	3.1688
Gruppe 2	3.1977	3.1130
Gruppe 3	4.4225	4.2507
Gruppe 4	4.4382	3.9367

Tabelle 6.4: Übersicht der Ergebnisse der gefahrenen Routen der Nutzer

Für die Prüfung zwischen welchen Studiengruppen ein signifikanter Unterschied vorliegt, wird der Dunn-Test [92] verwendet, welcher die Signifikanz der Unterschiede

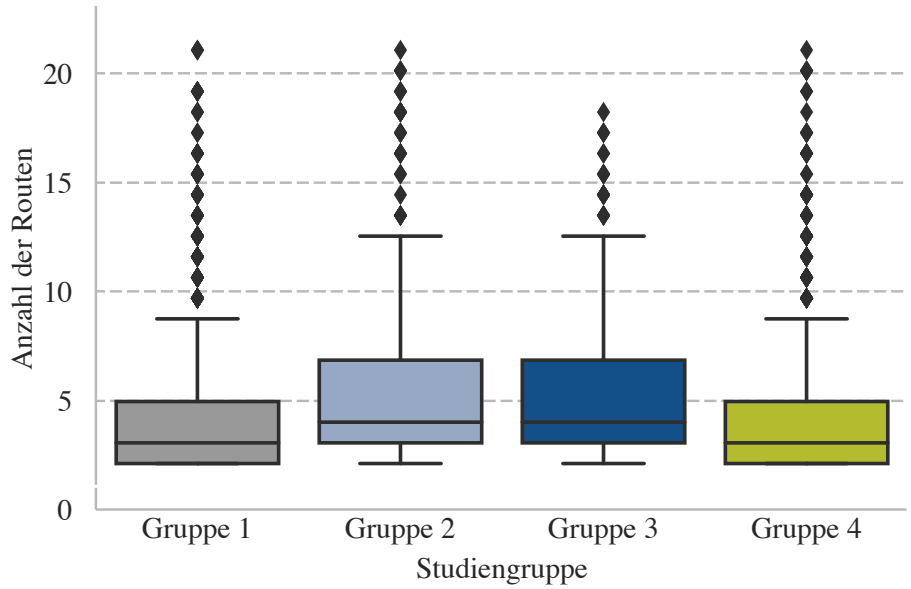


Abbildung 6.10: Anzahl der gefahrenen Routen für jede Studiengruppe

zwischen den Studiengruppen berechnet. Die Prüfung ergibt, dass jeweils ein signifikanter Effekt zwischen den Gruppen 3 und 4 gegenüber den Gruppen 1 und 2 vorliegt ($p_{13} = 4.746580e-09$, $p_{14} = 7.528428e-09$, $p_{23} = 1.977438e-08$ $p_{24} = 2.523727e-08$). Folglich kann abgeleitet werden, dass die Nutzer *NUNAV Navigation* häufiger verwenden, wenn sie *Context*-abhängige Erklärungen erhalten, unabhängig davon, ob diese kombiniert mit den statischen Erklärungen erfolgen. Folglich erfüllen nur die *Context*-abhängigen Erklärungen die Anforderung NFR1.

Die Hypothesen H2.1 und H2.2 müssen allerdings abgelehnt werden, da weder alle Erklärungstypen zu einer signifikant häufigeren Nutzung von NUNAV führen, noch das Bereitstellen beider vorgestellten Typen von Erklärungen zu einer erhöhten Nutzung pro Studienteilnehmer führt im Vergleich zur Präsentation von nur einem der integrierten Erklärungstypen.

Nutzerabweichungen von der vorgeschlagenen Route

Im Folgenden wird die Anforderung NFR2 geprüft, welche fordert, dass die Nutzer weniger häufig von der vorgeschlagenen Route abweichen sollen. Zur Prüfung werden ebenfalls zwei Hypothesen aufgestellt:

H3.1 WENN der Teilnehmer Erklärungen erhält, DANN folgt er signifikant häufiger der vorgeschlagenen Route als wenn er keine erhält.

H3.2 WENN der Teilnehmer nur einen der beiden vorgestellten Erklärungstypen erhält, DANN folgt er signifikant weniger häufig der vorgeschlagenen Route als wenn ihm beide Arten von Erklärungen präsentiert werden.

Als Messwert wird die Anzahl der Abweichungen relativ zu Gesamtlänge der Route verwendet. Die Einheit ist Abweichungen pro Kilometer. Zu diesem Messwert,

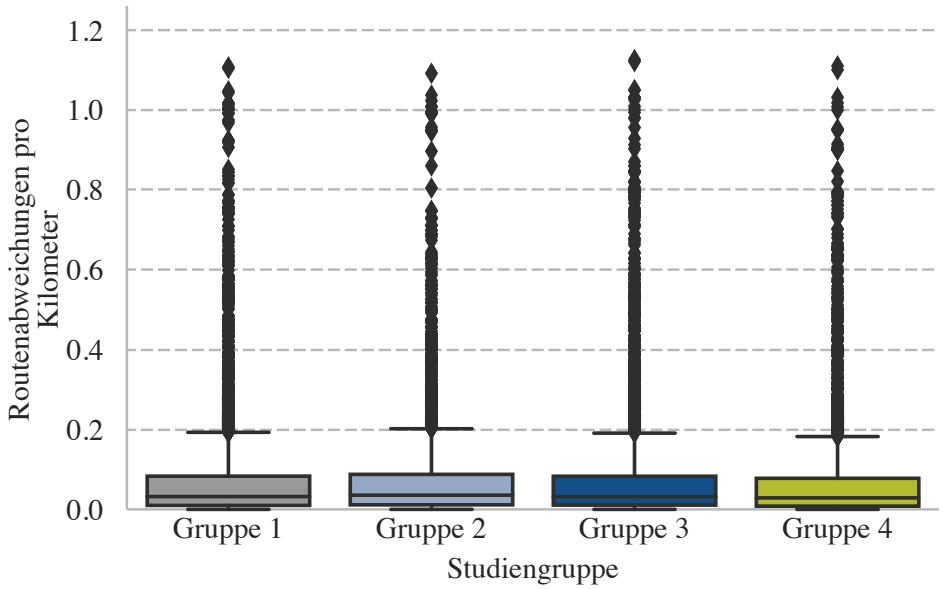


Abbildung 6.11: Anzahl der Abweichungen von der vorgeschlagenen Route pro Kilometer für jede Studiengruppe

muss erwähnt werden, dass die Erkennung, wann *End User* von der Route abweichen nicht trivial ist. Bei einer wirklichen Abweichung wird dies zum Teil mehrfach gezählt und es kann generell schnell zu falsch positiven Messungen kommen. Aufgrund der großen Datenmenge wird allerdings davon ausgegangen, dass diese Fehler einen vernachlässigbaren Effekt haben. Bei der Betrachtung von Abweichungen von der Route werden außerdem, wie zuvor beschrieben, Navigationen, bei denen es häufig zu einer ungenauen Positionierung kommt, aus dem Datensatz herausgefiltert. Außerdem werden auch hier Ausreißer in den Daten mit der gleichen Bedingung herausgefiltert, sodass schlussendlich 16 314 einzelne Navigationen zur Analyse der Routen-Abweichungen im Datensatz verblieben sind. Die Daten für jede Studiengruppe sind in Abbildung 6.11 dargestellt. Die berechneten Ergebnisse sind in Tabelle 6.5 zusammengefasst.

Studiengruppe	Mittelwert [N/km]	Standardabweichung [N/km]
Gruppe 1	0.0753	0.1293
Gruppe 2	0.0786	0.1272
Gruppe 3	0.0803	0.1410
Gruppe 4	0.0732	0.1328

Tabelle 6.5: Übersicht der Ergebnisse der Routen-Abweichungen pro Kilometer

Da auch die Daten der Routenabweichugen nicht normalverteilt sind (Shapiro-Wilk-Test: $p = 0.0$), wird zur Signifikanzprüfung der Kruskal-Wallis-Test verwendet [34]. Aufgrund des Ergebnisses von $p = 0.0000$, wird abgeleitet, dass ein Haupteffekt vorliegt. Analog zur vorherigen Analyse wird mit dem Dunn-Test die Signifikanz zwischen den einzelnen Studiengruppen geprüft.

Auch hier wird wieder von einem Signifikanzniveau $p < 0.05$ ausgegangen. Folglich weichen die Nutzer der Gruppe 4 signifikant weniger von der vorgeschlagenen Route ab, als die Teilnehmer aller anderen Gruppen ($p_{14} = 0.0217$, $p_{24} = 0.0000$, $p_{23} = 0.0029$). Weitere signifikante Unterschiede gibt es nicht.

Hypothese H3.1 trifft beim Bereitstellen aller Erklärungstypen (Gruppe 4) zu, kann jedoch für die Gruppen 2 und 3 nicht bestätigt werden. Folglich wird diese abgelehnt. Hypothese H3.2 kann angenommen werden, da die Nutzer der Gruppe 4 signifikant weniger von der vorgeschlagenen Route abgewichen sind als die *End User* der Gruppen 2 und 3.

Zusammenfassend kann die Anforderung NFR2 durch das Bereitstellen von allen Erklärungen erfüllt werden. Einzelne Erklärungen weisen keinen signifikant positiven Effekt auf die Anzahl der Abweichungen von der Route auf. Möglich ist folglich, dass die Effekte, der beiden Erklärungstypen nicht reichen. Eine genaue Begründung kann an dieser Stelle allerdings noch nicht gegeben werden.

Nutzerzufriedenheit mit der aktuellen Route

Um die Zufriedenheit der Nutzer mit der abgeschlossenen Navigation zu evaluieren, setzt NUNAV auf eine Bewertung mithilfe von 5 Sternen bei Erreichen des Zieles. Dies ist verknüpft mit der Frage „Wie hat dir die Fahrt gefallen?“ (siehe Abbildung C.1). Außerdem sind die Fahrtdauer und zurückgelegte Kilometer angegeben. Die folgenden konkreten Hypothesen werden für die Analyse abgeleitet:

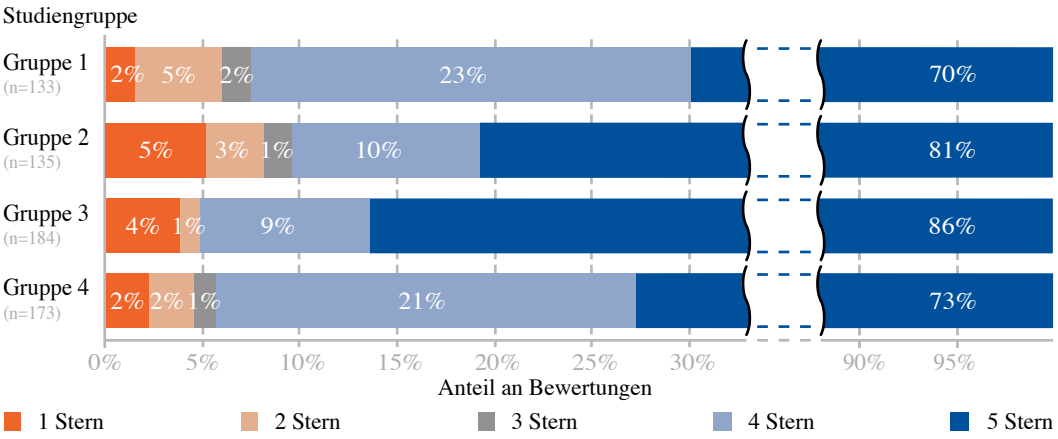


Abbildung 6.12: Verteilung der Teilnehmer-Bewertung der Navigation für jede Studiengruppe bezogen auf die insgesamt abgegebenen Bewertungen

H4.1 WENN Teilnehmer Erklärungen erhalten, DANN geben sie im Vergleich eine signifikant höhere Bewertung für die Navigation ab, als wenn sie keine Erklärungen erhalten.

H4.2 WENN Teilnehmer nur einen der beiden vorgestellten Erklärungstypen erhalten, DANN geben sie im Vergleich eine signifikant niedrigere Bewertung für die Navigation ab, als wenn sie beide Erklärungstypen erhalten.

Für die Prüfung zwischen welchen Studiengruppen ein signifikanter Unterschied vorliegt, wird ebenfalls der Dunn-Test [92] verwendet, nachdem ein Kruskal-Wallis-Test einen Haupteffekt zeigt ($p = 0.00335$). Daraus resultiert, dass es zwischen den Gruppen 1 und 3 ($p = 0.005723$) sowie 3 und 4 ($p = 0.024375$) einen signifikanten Unterschied bei der Zufriedenheit mit der aktuellen Route gibt.

Mithilfe von Abbildung 6.12 wird abgeleitet, dass es insbesondere mehr 5-Stern-Bewertungen in Gruppe 3 gegenüber den Gruppen 1 und 4 gibt. Die Anteile der ein und zwei Stern Bewertungen unterscheiden sich kaum. Folglich kann gesagt werden, dass die Teilnehmer signifikant zufriedener mit der Navigation sind, wenn sie Erklärungen wie in Gruppe 3 bekommen im Vergleich zu keinen Erklärungen (Gruppe 1) oder allen vorgestellten Erklärungstypen (Gruppe 4).

Da das Geben von Erklärungen die Nutzerzufriedenheit nicht in jedem Fall erhöht, muss Hypothese H4.1 abgelehnt werden. Hypothese H4.2 wird ebenfalls abgelehnt, da es unter anderem einen gegenteiligen Effekt zwischen den Gruppen 3 und 4 gibt.

Zusammenfassung des Einflusses auf die gesetzten Qualitätsziele

Bei der Analyse der Einflüsse der Erklärungen auf die gesetzten Qualitätsziele für *NUNAV Navigation* müssen die meisten Hypothesen abgelehnt werden. Signifikant negative Einflüsse durch Erklärungen konnten allerdings nicht gemessen werden. Der positive Effekt der integrierten Erklärungen fällt folglich geringer aus, als erwartet.

Ferner gibt es einen positiven Effekt durch Erklärungen auf die Nutzung pro Woche und Nutzer, wenn die *End User Context*-abhängige Erklärungen erhalten.

Sie nutzen NUNAV dann durchschnittlich 4,66 (*Context*-abhängig + statisch) bzw. 4,54 (*Context*-abhängig + statisch) mal pro Woche gegenüber 3,23 (statisch) bzw. 3,27 (keine Erklärung) mal pro Woche.

Die Anzahl der Routenabweichungen wird signifikant positiv beeinflusst (weniger Abweichungen), wenn die *End User* alle Erklärungen angezeigt bekommen (0,073 Abweichungen/km) gegenüber keinen Erklärungen (0,075 Abweichungen/km).

Außerdem geben die *End User* eine signifikant positivere Bewertung für die Route ab, wenn sie *Context*-abhängige Erklärungen erhalten (durchschnittlich 4,73 Sterne vs. 4,55 Sterne). Hier gibt es allerdings einen negativen Einfluss, wenn die *End User* zusätzlich die statischen Erklärungen erhalten, da diese die Routen dann signifikant schlechter bewerten (durchschnittlich 4,6 Sterne). Eine mögliche Erklärung wäre, dass sich die *End User* Fehler bei der Routenberechnung eher identifizieren können, wenn sie Erklärungen zum Algorithmus erhalten haben. Allerdings kann anhand der Daten keine sichere Begründung gegeben werden.

Folglich erfüllen die *Context*-abhängigen Erklärungen die Qualitätsanforderungen NFR1 sowie NFR3 und tragen somit zu einer signifikant häufigeren Nutzung von *NUNAV Navigation*. Zudem wird die Zufriedenheit der *End User* mit der Navigation signifikant gesteigert. Folglich kann empfohlen werden, diese weiterhin in der Anwendung zu integrieren.

Für die statischen Erklärungen allein kann kein signifikant positiver Effekt gemessen werden. Jedoch kann ein zusätzlicher, signifikant positiver Effekt im Rahmen von NFR2 gezeigt werden (weniger Routenabweichungen), wenn beide Erklärungstypen angezeigt werden. Allerdings hat diese Kombination der Erklärungen zugleich einen negativen Effekt auf die Routen-Zufriedenheit der *End User*, verglichen mit dem alleinigen Einsatz von *Context*-abhängigen Erklärungen. Für die statischen Erklärungen kann hier folglich keine klare Empfehlung abgeleitet werden.

6.2.5 Direkte Evaluation der Erklärungen

Die zuvor analysierten Metriken betreffen ausschließlich die Auswirkungen von Erklärungen auf die Qualitätsziele von *Graphmasters* für die Integration – im konkreten Fall *Usage Increase*, *System Acceptance* und *Satisfaction*. Um einen besseren Überblick über andere Qualitätsaspekte der Erklärungen zu bekommen, erfolgt zusätzlich eine direkte Evaluation der Erklärungen. Dies ist nötig, da weitere Ergebnisse benötigt werden, um die Resultate der *Case Study* zu interpretieren.

Ziel der direkten Evaluation

Mithilfe der Metriken der *Case Study* kann überprüft werden, welche Erklärungen, in welchen Kombinationen den zuvor aufgestellten Anforderungen genügen. Es kann allerdings für die statischen Erklärungen keine Empfehlung abgeleitet werden, da diese keine signifikanten Auswirkungen haben. Folglich ist die direkte Evaluation der Erklärungen nötig, wie es auch im Leitfaden vorgeschlagen wird. Insbesondere wird dort empfohlen nicht nur Verhaltensmetriken, sondern auch qualitative Evaluationen von *End Usern* zur Bewertung von Erklärungen zu nutzen. Daher ist im Anschluss an die *Case Study* ein Quasi-Experiment mit vier Nutzern durch durchgeführt worden. Diese zweite Analyse soll zusätzliche Daten liefern, um die Einflüsse der integrierten Erklärungen auf andere Qualitätsaspekte besser interpretieren zu können.

Das Ziel ist es folglich, konkrete Probleme oder Verbesserungspotenzial der entwickelten Erklärungen aufzudecken und somit die Effekte aus der *Case Study* zu erklären.

Methode

Die direkte Analyse der Erklärungen ist wiederum in zwei Teile gegliedert. Zunächst wurden während der *Case Study* einige Meta-Daten gesammelt. Es wurde beispielsweise aufgezeichnet, wie viele der *End User*, die die Möglichkeit hatten, auf die statischen Erklärungen zuzugreifen, diese genutzt haben und wie lange, sie diese im Fokus hatten.

Um qualitative Aussagen zu den Erklärungen zu erhalten wurde außerdem ein Quasi-Experiment durchgeführt. Bei diesem wurde vier Studienteilnehmern jeweils alle vier Erklärungen gezeigt. Darauf aufbauend sollten die Teilnehmer mit den gezeigten Erklärungen interagieren. Im Anschluss an jede Erklärung wurden ihnen Aussagen zur Bewertung auf einer Likert-Skala vorgelegt. Diese entsprechen den im Leitfaden unter Evaluation vorgestellten Aussagen (siehe Kapitel 6). Das bedeutet, mithilfe der Aussagen zu den Erklärungen werden Rückschlüsse auf die Qualitätsaspekte *Satisfaction*, *Perceived Transparency*, *Persuasiveness*, *Usefulness* und *Completeness* der Erklärungen gezogen. Darüber hinaus wurden Aussagen zum *Demand* der jeweiligen Erklärung hinzugefügt.

Außerdem sind die Teilnehmer gebeten worden, alle Gedanken beim Interagieren oder Lesen der Erklärungen mitzuteilen (*Think-Aloud*-Experiment). Anschließend an jede Erklärung haben alle Teilnehmer außerdem die Möglichkeit erhalten, Verbesserungsvorschläge für die Erklärungen zu machen oder auf fehlende Informationen in der Erklärung hinzuweisen. Der vollständige Fragebogen mit dem Ablauf des Quasi-Experiments ist in Abschnitt C zu finden.

Teilgenommen haben insgesamt drei Männer und eine Frau im Alter von 23 - 55 Jahren. Alle Nutzer waren mit Navigationsanwendungen wie Google oder Apple Maps vertraut und drei der vier Teilnehmer nutzen Navigationsanwendungen regelmäßig. Zwei der Teilnehmer haben außerdem sehr viel Erfahrung mit *NUNAV Navigation* (>10 mal verwendet), einem Teilnehmer war die Anwendung bekannt (2-3 mal verwendet) und eine Teilnehmerin hatte die Anwendung noch nie verwendet. Jeweils eine Person der Teilnehmergruppe passte zum Großteil auf eines der beiden vorgestellten Personas Unterabschnitt 6.1.3.

Im Folgenden werden die Ergebnisse der direkten Evaluation der Erklärungen vorgestellt.

Bedarf für die gegebenen Erklärungen

Die statischen Erklärungen (siehe Unterabschnitt 6.1.5), welche die Einflüsse auf den Routing-Algorithmus von NUNAV sowie das kollaborative Routing erklären, sind interaktiv. Somit kann evaluiert werden, wie viele Studienteilnehmer der *Case Study* diese wie häufig diese angefordert haben. Außerdem kann gemessen werden, wie lange sie die Erklärungen betrachtet haben. Die Studiengruppen, denen diese angezeigt wurden (Gruppe 2 und 4), enthalten insgesamt 1 766 Teilnehmer.

Als erfasst und nicht direkt wieder verlassen werden die Erklärungen gewertet, wenn die *End User* länger als 1,5 Sekunden in dem Dialog verbracht haben [93].

So können die Datensätze herausgefiltert werden, bei denen *End User* versehentlich den Dialog aufgerufen haben. Abbildung 6.13 zeigt die Häufigkeit der Aufrufe für die beiden statischen Erklärungstypen.

Um die Anzahl einzuordnen, wird betrachtet, ab welchem Nutzeranteil Funktionen durch *Graphmasters* in *NUNAV Navigation* integriert werden, worunter auch Erklärungen fallen. Anhand von internen Metriken ist abzuleiten, dass unter anderem mehrere Einstellungen von unter 5 % der *End User* genutzt werden. Folglich wird der Bedarf für Erklärungen ebenfalls ab einem Niveau von 5 % für gegeben angesehen. Zusätzlich zeigen Abbildung 6.14 und Abbildung 6.15 die Bewertung der Studienteilnehmer des Quasi-Experiments in Bezug auf deren Bedarf für die Erklärungen.

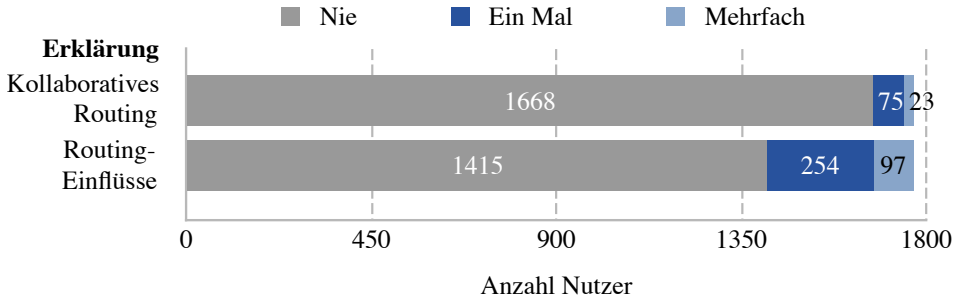


Abbildung 6.13: Häufigkeit der Aufrufe der Erklärungen zum Routing-Algorithmus sowie dessen äußeren Einflüssen

In der Abbildung 6.13 kann man erkennen, dass etwa 6 % der Nutzer auf die Erklärung zum kollaborativen Routing geklickt haben und folglich ein Bedarf besteht. Wie in Abbildung 6.14 zu erkennen ist, sagen drei von vier Teilnehmern des Quasi-Experiments, dass sie die Erklärung benötigt haben. Folglich herrscht eine Diskrepanz zwischen dem wirklichen Bedarf und den *End Usern* der *Case Study*, die die Erklärung angefordert haben. Insbesondere haben im Quasi-Experiment zwei der Teilnehmer angegeben, die Erklärung sich auch mehrfach anzusehen. Dennoch gegeben zwei Teilnehmer an, die beiden Erklärungstypen ausblenden können zu wollen. Dies ist anhand einer Teilnehmeraussage darauf zurückzuführen, dass die Informationen nach kurzer Zeit nicht mehr relevant sind, da das kollaborative Routing verstanden wurde.

Betrachtet man im Gegensatz dazu die gleichen Daten für die Erklärung zu den

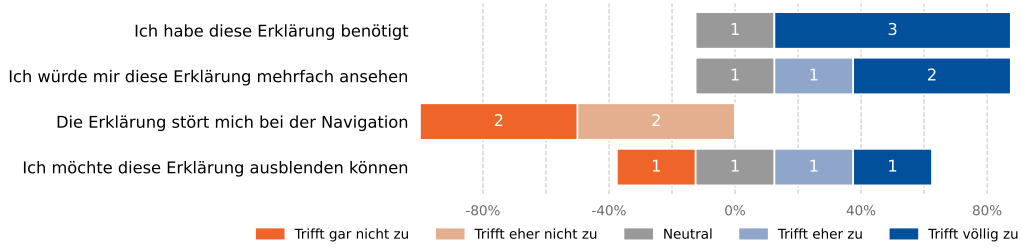


Abbildung 6.14: Subjektive Einschätzung des Bedarfs für die Erklärung zum kollaborativen Routing

Einflüssen auf die Routenberechnung fällt auf, dass etwa 20 % der Teilnehmer der *Case Study* sich die Erklärung angesehen haben. Wie in Abbildung 6.15 allerdings zu sehen ist, haben sowohl weniger Teilnehmer angegeben, dass sie die Erklärung benötigt haben, als auch sich mehrfach ansehen würden als bei der Erklärung zum kollaborativen Routing. Ein Erklärungsversuch kann durch die Aussage eines Teilnehmers am Quasi-Experiment getätigt werden. Der Teilnehmer hätte die Anzahl der Nutzer bereits verstanden und es war nicht eindeutig, dass eine weitere Erklärung aufrufbar ist. Außerdem hat eine Teilnehmerin hinzugefügt, dass die Möglichkeit zur Erklärung der Einflüsse des Routings zu gelangen neugieriger macht, als die Nutzerzahl. Der Button zum Aufruf enthält die Frage: „Wie ist meine Route entstanden?“ (siehe Unterabschnitt 6.1.5).

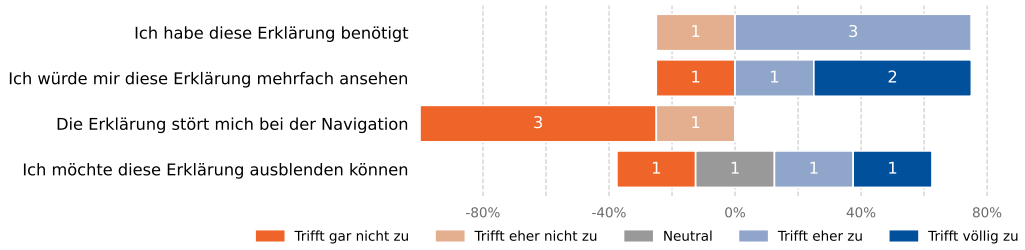


Abbildung 6.15: Subjektive Einschätzung des Bedarfs für die Erklärung zu Einflüssen auf den Routing-Algorithmus

Aus diesen freien Aussagen und dem Gegensatz zwischen den benötigten Erklärungen und den wirklich angeforderten, kann folglich abgeleitet werden, dass die Erklärung zum kollaborativen Routing offensichtlicher zu erreichen sein sollte. Die Idee eines weiteren Teilnehmers war, dass es möglich sein sollte, diese Erklärung über den Dialog „Trete Schwarm bei...“ erreichen zu können. Dies ist der Dialog, der zu sehen ist, während *NUNAV Navigation* die erste Route während der Nutzung lädt. Er hat außerdem hinzugefügt, dass an dieser Stelle der Schwarmbegriff unverständlich sei und durch die Erklärung klar wird.

Außerdem kann es dadurch, dass nur bis zu 20 % der *End User* in der *Case Study* die Erklärungen gelesen haben, sein, dass die Auswirkungen auf andere Qualitätsmerkmale in der Studie nicht im signifikanten Bereich lagen. Zumindest für die Erklärung zum kollaborativen Routing ließe sich dies anhand der Aussagen der Teilnehmer des Quasi-Experiments durch einen verbesserten Weg, um zur Erklärung zu gelangen, ggf. steigern. Daher sollte eine andere Variante in einer zweiten Iteration getestet werden. Des Weiteren gab es in einem Fall beim Quasi-Experiment Kritik daran, dass die Hilfe-Center-Artikel im Browser des Smartphones geöffnet würden. Eine Integration in der App würde der Aussage nach eher dazu bewegen, sich eine Erklärung auch durchzulesen.

Für die beiden *Context*-abhängigen Erklärungen gab es keine Metriken, welchen den Bedarf innerhalb der *Case Study* widerspiegeln können. Auch gab es im Rahmen des Quasi-Experiments lediglich positive Rückmeldungen zum Bedarf der Erklärungen. Des Weiteren wurden beide Erklärungen weder als störend empfunden, noch gab es Studienteilnehmer, die den Wunsch äußerten, die Erklärungen ausblenden zu wollen (siehe Abbildung 6.16 und Abbildung 6.17). Außerdem gab es auch

keine weiteren Äußerungen der Teilnehmer des Quasi-Experiments zu den beiden Erklärungen. Daher können aus den Daten zum Bedarf für diese Erklärungen keine Verbesserungsmöglichkeiten abgeleitet werden. Dies deckt sich mit den gezeigten positiven Einflüssen der Erklärungen in der *Case Study*.

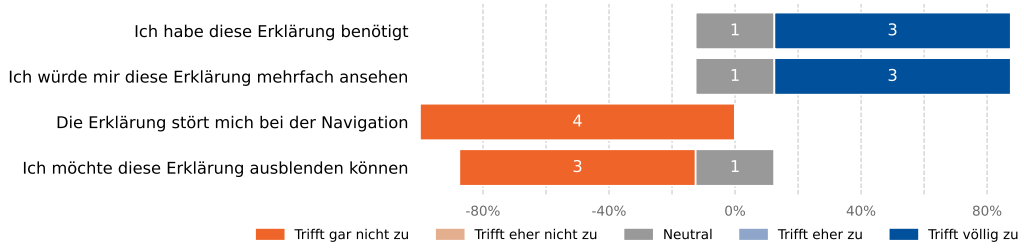


Abbildung 6.16: Subjektive Einschätzung des Bedarfs für die Erklärung zum aktuellen Verkehrsgeschehen

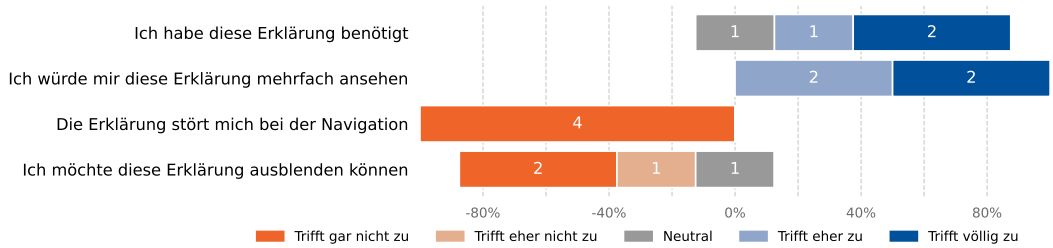


Abbildung 6.17: Subjektive Einschätzung des Bedarfs für die Erklärung zu Positionsungenauigkeiten

Inhalt der gegebenen Erklärungen

Um den Inhalt der entwickelten Erklärungen zu bewerten, kann vor allem die *Usefulness*, die *Perceived Transparency* und die *Completeness* der Erklärungen analysiert werden.

Für die beiden statischen Erklärungen kann indirekt die *Usefulness* der Erklärungen im Rahmen der *Case Study* bestimmt werden. Das Hilfe-Center, welches zur Anzeige der Erklärungen genutzt wurde, enthält am Ende jedes Artikels die Möglichkeit, diesen als „Hilfreich“ oder „Nicht hilfreich“ zu bewerten. Auf den Hilfe-Center können *End User* zwar auch über andere Wege, als die in der *Case Study* bereitgestellten gelangen. Da es hier allerdings um die Bewertung der Erklärung an sich geht, können die Daten hier dennoch verwendet werden. Tabelle 6.6 zeigt die Bewertungen als „Hilfreich“ oder „Nicht hilfreich“ für die beiden statischen Erklärungen an.

An den Bewertungen, welche aus dem Hilfe-Center stammen, kann man erkennen, dass ein Großteil (90 % kollaboratives Routing, 87 % Routing-Einflüsse) der *End User*, welche die Artikel bewertet haben, diese als „hilfreich“ bewertet haben. Dies spiegelt sich auch in der subjektiven Bewertung der Erklärungen im

Artikel	Hilfreich	Nicht Hilfreich
Kollaboratives Routing	76	9
Einflüsse auf die Routenberechnung	209	32

Tabelle 6.6: *Usefulness* der Hilfe-Center-Artikel

Quasi-Experiment wider. Mit einer Ausnahme stehen die Studienteilnehmer den Aussagen, dass die Erklärungen nützlich sind und diese verstanden haben, positiv oder neutral gegenüber (siehe Abbildung 6.18 und Abbildung 6.19). Lediglich ein Teilnehmer hat die Aussage, dass die Erklärung zum kollaborativen Routing nützlich für die Navigation ist mit „Trifft eher nicht zu“ bewertet. Auf Nachfrage, liegt dies daran, dass der Teilnehmer sich wünschen würde zu sehen, wann das kollaborative Routing ihn direkt während der Navigation beeinflusst. Folglich wird eine weitere *Context*-abhängige Erklärung zusätzlich zur Anzahl der *End User* in der Umgebung benötigt, welche diesen Umstand stärker herausstellt. Die Umsetzung ist aufgrund der benötigten Daten allerdings schwierig.

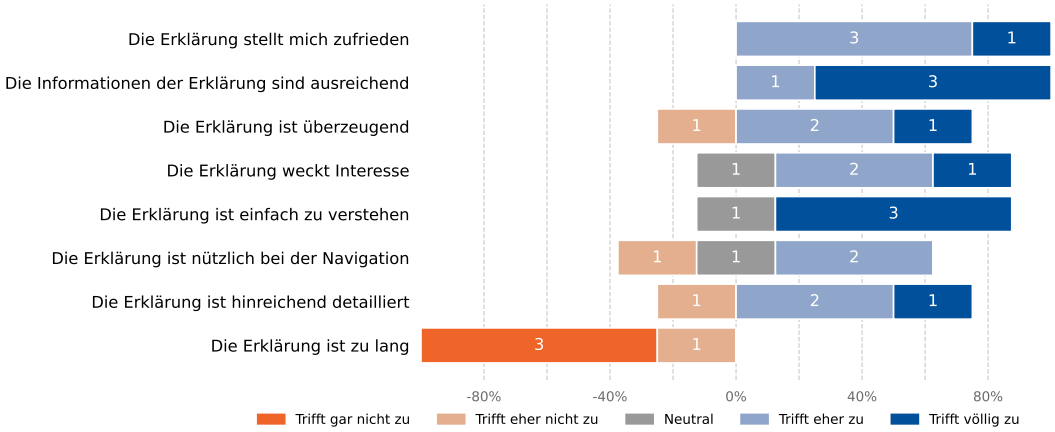


Abbildung 6.18: Subjektive Einschätzung des Inhalts für die Erklärung zum kollaborativen Routing

Bei der Betrachtung von *Perceived Transparency* und *Completeness* sind die Bewertungen für die statischen Erklärungen im Quasi-Experiment überwiegend positiv oder neutral (siehe Abbildung 6.18). Eine Aussage zur Erklärung zum kollaborativen Routing, die wiederum negativ beantwortet wurde, stammt von selbigem Nutzer wie zuvor und ist auch ähnlich begründet. Außerdem wurde die Aussage zur Länge der Erklärung zu den Einflüssen auf die Routenberechnung in einem Fall als zu lang bewertet. Die Teilnehmerin hätte sich Formulierungen gewünscht, die etwas schneller das Kernthema erfassen.

Ein weiterer Wunsch eines Teilnehmers war, sehen zu können aus welchen Quellen die Daten kommen, die einen Einfluss auf das Routing haben, um deren Güte zu bewerten. Zusammenfassend kann für den Inhalt der statischen Erklärungen gesagt werden, dass dieser im Allgemeinen wenig Kritikpunkte im Quasi-Experiment erfahren hat und lediglich an einzelnen Stellen verbesserungswürdig ist (z.B. in

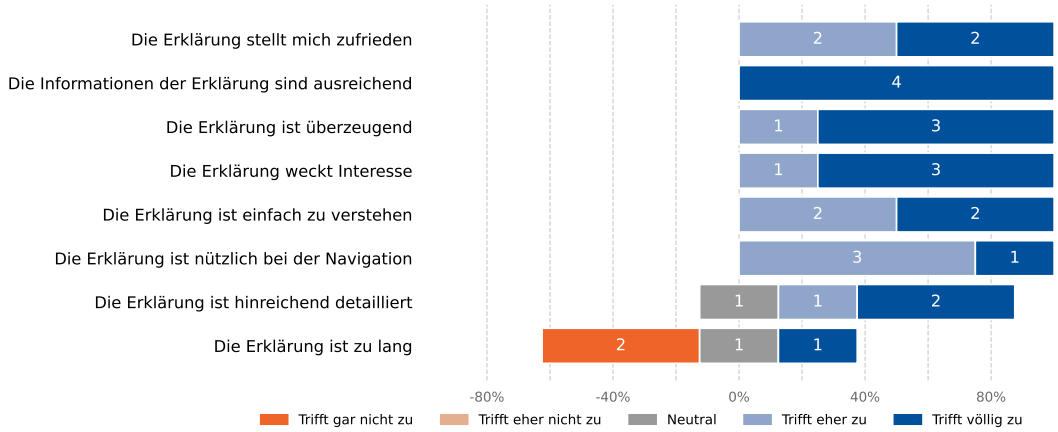


Abbildung 6.19: Subjektive Einschätzung des Inhalts für die Erklärung zu Einflüssen auf den Routing-Algorithmus

den Formulierungen). Die Forderung nach der Preisgabe der Datenquellen für die Einflüsse auf das Routing ist aufgrund von Verträgen zum Großteil nicht möglich.

Da die Bewertungen der Aussagen zu den beiden *Context*-abhängigen Erklärungen deutlich unterschiedlich ausfallen, werden diese getrennt behandelt. Für die Erklärung zum aktuellen Verkehrsgeschehen kann gesagt werden, dass der Inhalt ausschließlich positiv bewertet wurde (siehe Abbildung 6.20). Ferner sind zu den Bewertungen der vorgegebenen Aussagen weitere positive Kommentare hinterlassen worden. Zum Beispiel haben zwei Teilnehmer bereits positive Erfahrungen mit ähnlichen Funktionen von Konkurrenzprodukten gemacht und die Erklärung aufgrund ihrer Bekanntheit begrüßt. Folglich muss an der Erklärung an sich aufbauend auf der vorliegenden Analyse nicht verändert werden.

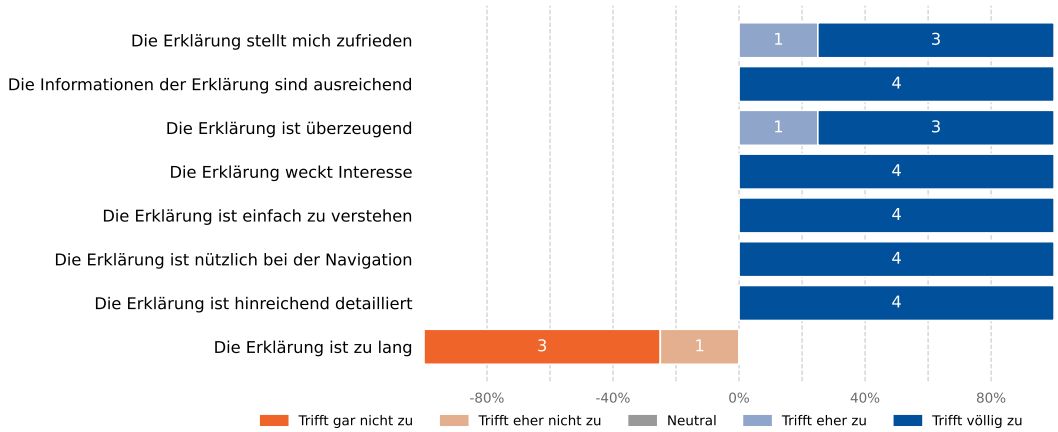


Abbildung 6.20: Subjektive Einschätzung des Inhalts für die Erklärung zum aktuellen Verkehrsgeschehen

Die Erklärung zur ungenauen Positionierung hat, wie in Abbildung 6.21 zu erkennen ist, einzelne negative Bewertungen bezüglich der *Satisfaction* und *Perceived Transparency*. Dies ist darauf zurückzuführen, dass zwei Teilnehmer versucht hatten, über das *Growl*, der zur Erklärung gehört, eine genaue Erklärung der Funktion anzufordern und diese nicht erhalten konnten. Dies wurde mit der Frage danach verbunden, was insbesondere „Position ungenau“ bedeutet. Ein Vorschlag zur Lösung war, dass jedes „Growl“ wie bei der Erklärung zum kollaborativen Routing eine kurze sowie eine lange Erklärung erhält. Die Interaktionsart ist laut einem Kommentar innerhalb des Quasi-Experiments die Erwartung dadurch, dass sie an einer Stelle möglich ist. Folglich sollte in einer weiteren Iteration aufgrund der Konsistenz dieser Vorschlag umgesetzt werden.

Gestaltung der Erklärungen

Neben der durchgeführten Analyse der verschiedenen Qualitätsmerkmale gab es außerdem vereinzelt Rückmeldungen zur *Presentation* der Erklärungen.

Für die Erklärung des kollaborativen Routings wurde vorgeschlagen, dass die Erklärung mehr mit Grafiken arbeiten sollte, um die Funktionsweise des Algorithmus zu erklären.

Positiv wurden die Sprachansagen, welche die Erklärung zum Verkehrsfluss und zur ungenauen Positionierung begleiten, von mehreren Teilnehmern des Quasi-Experiments hervorgehoben. Diese seien explizit bei der Navigation im Auto sehr hilfreich, da man nicht immer auf das Smartphone sehen kann.

6.2.6 Validität der Evaluation von Erklärungen

Im Folgenden wird untersucht, inwiefern die Ergebnisse der durchgeführten *Case Study* und des Quasi-Experiments eingeschränkt sein könnten. Für die beiden Evaluation werden die von Wohlin et al. vorgestellten *Threads to Validity* überprüft [34].

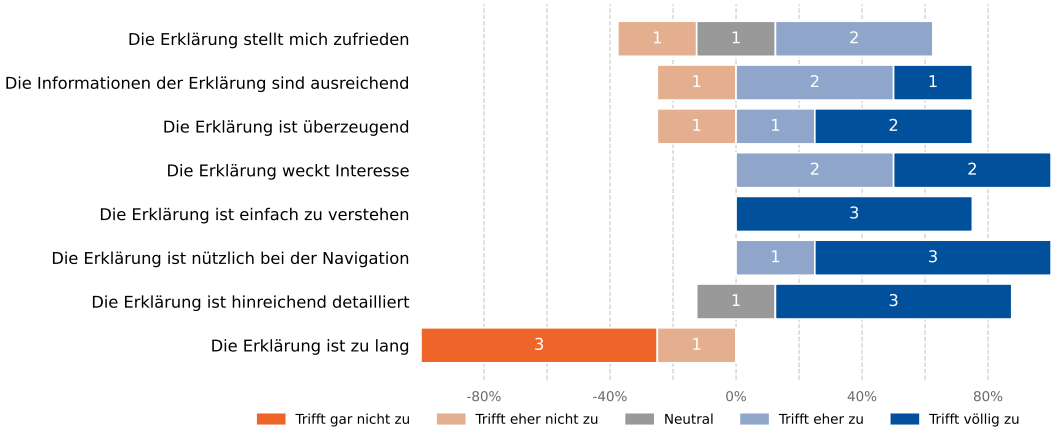


Abbildung 6.21: Subjektive Einschätzung des Inhalts für die Erklärung zu Positionsungenauigkeiten

Die *Conclusion Validity* bezieht sich auf die statistische Analyse der Ergebnisse und die Zusammensetzung der Probanden. Unter *Internal Validity* werden Einflüsse zusammengefasst, die die unabhängige Variable in Bezug auf die Kausalität ohne die Kenntnis der analysierenden Personen beeinflussen können. Die *External Validity* ist die Verallgemeinerbarkeit von Ergebnissen auf andere Kontexte und *Construct Validity* behandelt die Möglichkeit der Verallgemeinerung von Versuchsergebnissen auf ein Konzept oder die Theorie hinter dem Versuch.

In Bezug auf die statistische Analyse der *Case Study* konnte im Vergleich zu vergangenen Arbeiten eine große Datenmenge analysiert werden ($\approx$ 4000 Teilnehmer). Außerdem waren die *Teilnehmer* der Studie zufällig auf die verschiedenen Bedingungen verteilt. Folglich wird abgeleitet, dass eine hohe *Conclusion Validity* vorliegt. Die Daten wurden mithilfe von Hypothesentests und einem Signifikanzniveau von 5% analysiert.

Da bei der *Case-Study* die meisten äußeren Einflüsse nicht überprüfbar oder unbekannt sind, gibt es Einschränkungen für die *Internal Validity*. Allerdings konnten auf Basis von Einflusshypothesen außerhalb der Evaluation von Erklärungen bestimmte Effekte ausgeschlossen und nicht verwendbare Daten herausgefiltert werden.

Die *Conclusion Validity* des *Quasi-Experiments* ist sehr gering, da weder eine zufällige Verteilung der Teilnehmer auf mehrere Bedingungen oder eine verschiedene Zusammensetzung der Bedingungen durchgeführt wurde, noch bei vier Teilnehmern statistisch signifikante Aussagen getroffen werden können. Allerdings war dies nicht das Ziel dieser Evaluation. Das *Quasi-Experiment* ist für die bessere Interpretation der Daten der *Case Study* durchgeführt worden.

Für die *Internal Validity* wurden für alle Teilnehmer die gleichen Umstände geschaffen. Das verschiedene Vorwissen zu *NUNAV Navigation* der Probanden kann die Ergebnisse beeinflussen. Da das Vorwissen allerdings ebenfalls gemessen wurde, ist dies in die Interpretation der Ergebnisse mit eingeflossen.

Die zwei verbleibenden *Threads to Validity* werden für die beiden Studien zusammen betrachtet. Die Qualität der Erklärungen, die aus den beiden Evaluationen hervorgeht, kann nur für den konkreten Anwendungsfall der Navigation, verallgemeinert werden (*External Validity*). Allerdings muss bei der Verallgemeinerung der Erklärungen zum Routing-Algorithmus einschränkend erwähnt werden, dass diese sehr spezifisch für den *NUNAV*-Routing-Algorithmus sind.

Da beide Evaluationen zu dem Zweck durchgeführt wurden, die Anwendbarkeit des entwickelten Leitfadens zu prüfen, wird die Auswirkung auf den Leitfaden betrachtet. Hier kann anhand der Studien nur die Anwendbarkeit für die *Graphmasters GmbH* eindeutig bewertet werden (*Construct Validity*).

Zusammenfassend kann gesagt werden, dass die *Conclusion Validity* und *Internal Validity* der beiden Evaluationen gegensätzlich bewertet wurden und somit verschiedene Ergebnisse liefern, welche zusammengefügt werden können. Folglich hat das ergänzende Quasi-Experiment, die Aufgabe des Ausgleiches der fehlenden Interpretationsmöglichkeiten der *Case Study* erfüllt.

Die Einschränkung der *External Validity* auf den Kontext Navigation und zum Teil den konkreten Routingalgorithmus von *Graphmasters* wurde bewusst gewählt. Die Anwendung des Leitfadens sollte zeigen, ob dieser in einem konkreten Kontext

anwendbar ist und war nicht darauf ausgerichtet, die Ergebnisse für die Qualität von bestimmten Erklärungen verallgemeinern zu können.

Für die *Construct Validity* kann abgeleitet werden, dass mithilfe des Leitfadens im agilen Prozess von *Graphmasters* Erklärungen mit positiven Einflüssen auf Softwarequalität entwickelt wurden. Für welche weiteren Prozesse der entwickelte Leitfaden geeignet ist, muss allerdings noch gezeigt werden.

6.2.7 Zusammenfassung der Evaluation

Anhand einer *Case Study* innerhalb der Produktivversion von *NUNAV Navigation* wurden zunächst die Qualitätsanforderungen von *Graphmasters* auf die Einflüsse der integrierten Erklärungen untersucht. Diese Prüfung hat ergeben, dass die Erklärung zum kollaborativen Routing sowie zu den Einflüssen auf den Routing-Algorithmus keinen signifikanten Effekt auf die Metriken der aufgestellten Anforderungen aufweisen (siehe Unterabschnitt 6.1.4).

Als positives Ergebnis der *Case Study* kann hervorgehoben werden, dass die Erklärungen zum aktuellen Verkehrsgeschehen und zu Positionsungenauigkeiten während der Navigation einen positiven Einfluss auf die *Route Satisfaction* der *End User* haben. Des Weiteren konnte gezeigt werden, dass *End User*, die diesen Erklärungstyp statt der statischen Erklärungen erhalten, *NUNAV Navigation* im Durchschnitt häufiger pro Woche zur Navigation verwenden.

Außerdem zeigt die Analyse der *Case Study*, dass das Bereitstellen aller Erklärungstypen einen signifikant positiven Effekt auf die *Route Acceptance* hat. Im Vergleich zum Bereitstellen der *Context*-abhängigen Erklärungen ist allerdings ein signifikant negativer Einfluss auf die Route Satisfaction messbar. Für letzteres konnte in keiner der beiden durchgeführten Studien ein Grund ermittelt werden.

Aufgrund der mangelnden Ergebnisse zu den Eigenschaften der entwickelten Erklärungen wurde im Anschluss ein Quasi-Experiment durchgeführt, welches zusammen mit weiteren Metadaten aus der *Case Study* für die beiden statischen Erklärungen potenzielle Verbesserungen herausgestellt hat.

Für die *Context*-abhängigen Erklärungen konnten keine konkreten Vorschläge oder Ergebnisse aus dem Quasi-Experiment gefolgert werden. Dies hat allerdings die Qualität der Erklärungen, welche aus der *Case Study* gefolgert wurde, bestätigt.

Als Konsequenz sollten die in Abschnitt 6.2.5 erfolgten Vorschläge der Teilnehmer des *Quasi-Experiments* in eine zweite Iteration der Erklärungen integriert werden. Durch eine weitere Evaluation kann im Anschluss festgestellt werden, ob diese wie vermutet, zu einer erhöhten Qualität der evaluierten Qualitätsaspekte führt.

Kapitel 7

Diskussion

Ziel der vorliegenden Arbeit war die Entwicklung eines Modells respektive Richtlinien zur Unterstützung bei der Gestaltung von Erklärungen. Innerhalb dieser Arbeit ist ein Leitfaden entstanden, welcher die geforderte Unterstützung liefern soll. In diesem Kapitel wird die Anwendbarkeit des Leitfadens in der Wirtschaft analysiert. Zudem wird die Allgemeingültigkeit des enthaltenen Modells und des Katalogs der Zusammenhänge diskutiert, sowie die Limitierungen der Evaluation der Erklärungen aufgezeigt. Außerdem werden die Herausforderungen dieser Arbeit vorgestellt.

7.1 Interpretation der Ergebnisse

Für die Einordnung der Ergebnisse aus den beiden Teilen dieser Arbeit ist vorwiegend interessant, inwiefern der im ersten Teil entwickelte Leitfaden in der Praxis anwendbar ist. Dabei werden zunächst die enthaltenen Ergebnisse zusammengefasst. Für die Beurteilung des Nutzens des enthaltenen Modells, der Auswirkungen auf Qualitätsaspekte und der Heuristiken zur Integration von Erklärungen in ein bestehendes System wurde der Leitfaden bei der Firma *Graphmasters* eingesetzt. Um die Anwendbarkeit des entwickelten Leitfadens zu analysieren, werden im Folgenden zum einen die Erfahrungen, die während der Anwendung gemacht wurden, ausgewertet. Zum anderen wird die Qualität der im Rahmen der Anwendung entstandenen Erklärungen betrachtet.

Der Leitfaden ist im Rahmen einer Literaturrecherche entstanden, in welcher die bestehenden Ergebnisse für das Design von Erklärungen in erklärbaren Systemen analysiert wurden. Das im Leitfaden für die Integration von Erklärungen enthaltene Modell beinhaltet die folgenden Teile:

RQ1 Unter *External Dependencies* sind alle Rahmenbedingungen zusammengefasst, die einen direkten Einfluss auf die Anforderungen an Erklärungen aufweisen. Als relevante Aspekte sind verschiedene Ausprägungen des *Contexts* von erklärbaren Systemen, sowie die *Objectives* für die Integration von Erklärungen auf verschiedenen Abstraktionsebenen in dem Modell enthalten. Somit unterstützt dieser Modellteil die Anforderungserhebung für Erklärungen.

RQ2 Die Umsetzung der Anforderungen wird in dem vorgestellten Modell durch in der Forschung evaluierte *Characteristics* unterstützt. Dabei enthält das Modell Eigenschaften von Erklärungen für die Umsetzung des Bedarfs für Erklärungen (*Demand*), die transportierten Informationen (*Content*) und die Art der Informationsvermittlung an *End User* (*Presentation*). Die im Modell vorgestellten Ausprägungen bieten unterschiedliche Möglichkeiten, um Erklärungen für verschiedene Kontexte zu gestalten. Aufgeführt sind lediglich Eigenschaften, für die ein Effekt auf die Qualität von Softwaresystemen bereits gezeigt werden konnte.

RQ3 Über die Entwicklung von Erklärungen hinaus bietet das Modell Hilfestellungen für die Evaluation von Erklärungen in einem System. Grundsätzlich werden die im *Software Engineering* üblichen Studienformen vorgeschlagen, welche je nach Ziel einer Evaluation Anwendung finden können [vgl. 34]. Außerdem wird zwischen der direkten Messung der Qualität von Erklärungen sowie der Messung der Einflüsse von integrierten Erklärungen auf externe Qualitätsaspekte eines Systems unterschieden.

Um einen ganzheitlichen Überblick über die Qualität von Erklärungen zu erhalten, empfiehlt der Leitfaden, in den das Modell integriert ist, eine Kombination der verschiedenen Evaluationsmöglichkeiten. Die Erklärungen können sowohl direkt evaluiert werden als auch deren Einflüsse auf andere Qualitätsaspekte gemessen werden. Für die Bewertung der Qualität sind außerdem sowohl quantitative als auch qualitative Metriken notwendig, um die Performanz der *End User* wie auch deren subjektive Wahrnehmung zu betrachten.

RQ4 Neben dem Modell für Erklärungen, enthält der Leitfaden einen Katalog der Erklärungen, welcher die Zusammenhänge der Eigenschaften von Erklärungen auf abhängige Qualitätsaspekte zusammenfasst. Dies beantwortet zum einen die Frage, unter welchen Bedingungen das Präsentieren von Erklärungen, welche Einflüsse auf ausgewählte externer Softwarequalitätsaspekte hat. Zum anderen werden ebenfalls die Auswirkungen der Granulariät von Erklärungen auf selbige Qualitätsaspekte zusammengefasst.

Abgeleitet aus den ersten Teilen des Leitfadens werden des Weiteren Heuristiken für die Gestaltung von Erklärungen. Dabei werden allgemein anwendbare Zusammenhänge vorgeschlagen, welche im Regelfall für die Integration von Erklärungen beachtet werden sollten.

Die im entwickelten Leitfaden enthaltenen Ergebnisse sind nicht als vollständiger Überblick, über die Möglichkeiten der Gestaltung von Erklärungen zu sehen. Viel mehr soll der Leitfaden eine Unterstützung und Anregungen zur Integration von Erklärungen bieten. Diese Empfehlung geschieht auf Basis von existierenden Ergebnissen, wodurch eine Orientierung an den Ergebnissen im Leitfaden in der Regel zu positiven Resultaten bei der Integration von Ergebnissen führen kann. Welche Einschränkungen für diese Aussage gelten, muss allerdings zunächst in zukünftigen Arbeiten überprüft werden.

Als erste Prüfung der Aussage wurde der Leitfaden im Rahmen dieser Arbeit initial in der Firma *Graphmasters* eingesetzt. Der Einstieg in die Anwendung des

Leitfadens ist innerhalb eines Workshops erfolgt, welcher zusätzlich zur Vorstellung des Leitfadens mit einer Einführung in das Thema Erklärbarkeit begonnen hat (siehe Unterabschnitt 6.1.2). Anhand der Rückfragen, zum Beispiel, wo die genaue Abgrenzung zwischen Erklärbarkeit und gutem User-Interface-Design liegt, wurde deutlich, dass eine Anwendung des Leitfadens nicht ohne Vorwissen zum Thema *Explainability* erfolgen kann. Der Leitfaden an sich enthält diese Einführung allerdings nicht.

Anhand von weiteren Rückfragen bei der Vorstellung des Leitfadens zur Integration von Erklärungen konnten außerdem Verständnisprobleme innerhalb des Leitfadens identifiziert werden. Insbesondere die Abgrenzung zwischen verschiedenen Inhaltstypen von Erklärungen haben nicht alle Teilnehmer des Workshops direkt verstanden. Die Typen waren zum Zeitpunkt der Durchführung als Fragewörter voneinander abgegrenzt. Für das finale Modell wurden daher einzelne Begriffe des Modells überarbeitet (siehe Abbildung 5.1). Mit diesen integrierten Änderungen wird darauf geschlossen, dass Nutzer des Leitfadens mit Vorwissen über das Thema Erklärbarkeit diesen zu großen Teilen verstehen können. Eine direkte Evaluation der Verständlichkeit des Leitfadens ist allerdings im Rahmen dieser Arbeit nicht erfolgt.

Außerdem war während des Workshops zu beobachten, dass insbesondere die verschiedenen Möglichkeiten im Leitfaden, Erklärungen zu gestalten, die Diskussion zu Umsetzungsideen angeregt haben. Somit haben sich die verschiedenen *Characteristics* zusammen mit den Einflüssen auf verschiedene Qualitätsaspekte als sehr hilfreich für die Entwicklung von Erklärungen herausgestellt. Zusätzlich hat das Modell geholfen, die Ideen zur Evaluation zu systematisieren und folglich Metriken festzulegen. Dabei konnte festgestellt werden, dass nicht alle Teile des Modells in jedem Kontext anwendbar sind.

Außerdem konnte der Leitfaden aus eigener Erfahrung bei der Konkretisierung der Ergebnisse des Workshops helfen. Die *Objectives* im Modell boten dabei eine gute Hilfestellung zum Aufstellen eines Qualitätsmodells. Der Katalog der Zusammenhänge von bestimmten Eigenschaften, von Erklärungen und die Heuristiken für die Gestaltung haben die Umsetzung des Designs gut unterstützt.

Aus der Evaluation der entstandenen Erklärungen kann des Weiteren gefolgert werden, dass durch die Anwendung des Leitfadens Erklärungen entwickelt werden konnten, welche positive Einflüsse auf die Softwarequalität eines Systems haben. Die Untersuchung des Einflusses der Erklärungen hat aber auch gezeigt, dass nicht alle entwickelten Erklärungen einen positiven Einfluss auf die Softwarequalität erreichen konnten. Hier hat allerdings die anhand des Leitfadens entwickelte zusätzliche direkte Evaluation der Erklärungen mögliche Probleme und Verbesserungsmöglichkeiten aufgedeckt. Diese können in einer weiteren Iteration der Erklärungen umgesetzt werden. Ferner deuten weder qualitative noch quantitative Ergebnisse der Evaluation auf negative Effekte durch die Erklärungen hin. Daher können diese ohne Bedenken über die Studie hinaus in *NUNAV Navigation* integriert bleiben.

Aus den Ergebnissen des Quasi-Experiments ist außerdem hervorgegangen, dass alle vier verschiedenen Erklärungen unterschiedlich von den Teilnehmern bewertet wurden. In der *Case Study* wurden die Erklärungen allerdings in zwei Gruppen zusammengefasst, um eine möglichst hohe Zahl an *End Usern* in jeder Studiengruppe innerhalb eines kurzen Zeitraums erreichen zu können. In einer erneuten Studie sollte folglich die Dauer verlängert und die Erklärungen einzeln

evaluiert werden. Auch könnte über einen längeren Zeitraum der Nutzerzuwachs von *NUNAV Navigation* zwischen den Studiengruppen verglichen werden. Dies war während der durchgeführten Studie nicht möglich.

Folglich wird abgeleitet, dass in dieser Arbeit ein Leitfaden entstanden ist, welcher die geforderte Unterstützung bei der Integration von Erklärung bietet. Dabei wurden bei der Anwendung des Leitfadens alle Teile des Modells sowie der Katalog der Zusammenhänge von Aspekten von Erklärungen angewendet. Zudem haben die Heuristiken im Leitfaden das konkrete Design unterstützt.

Weitere positive Ergebnisse, welche die Anwendung des Leitfadens in der *Graphmasters GmbH* mit sich gebracht hat, sind die entstandene Sensibilität für den Bedarf von Erklärungen bei *End Usern* und das entwickelte *Feature-Flag*-System, welches es ermöglicht in Zukunft einfacher Evaluationen für neu entwickelte Funktionen der mobilen Anwendungen durchzuführen.

7.2 Limitierungen des Leitfadens für Erklärungen

Der in dieser Arbeit entwickelte Leitfaden für die Integration von Erklärungen in Softwaresysteme fasst bereits gezeigte Ergebnisse zusammen und strukturiert diese. Folglich sind in dem entstandenen Modell für Erklärungen, dem Katalog über die Einflüsse von Erklärungen auf bestimmte Qualitätsaspekte und den daraus abgeleiteten Design Heuristiken keine neuen Ergebnisse entstanden, sondern nur bestehende Resultate in einen Zusammenhang gestellt worden. Der Leitfaden erhebt allerdings keinen Anspruch auf Vollständigkeit. Vor allem für spezifische Kontexte bieten bereits entwickelte Modelle einen tieferen Einblick [17, 37].

Eine weitere Limitierung, die bei der Entwicklung des Leitfadens auf Basis einer Literaturrecherche gemacht werden muss, ist, dass diese nur von einer Person durchgeführt und daher nicht überprüft wurde. Daher können Aspekte übersehen oder fehlerhaft eingeordnet worden sein. Dies unterscheidet die im Rahmen dieser Arbeit durchgeführte Literaturrecherche von bereits durchgeführten systematischen Literaturrecherchen zum Thema *Explainability* [vgl. 17, 5].

In dieser Arbeit wurde versucht, den Leitfaden ganzheitlich als Methode in der Wirtschaft anzuwenden. Dabei hat sich herausgestellt, dass nicht alle Teile des Leitfadens für jedes Team einsetzbar sind. Beispielsweise wurden die Ziele zur Integration von Erklärungen von *Graphmasters* sehr allgemein gehalten, da eine Konkretisierung der Ziele zum Beispiel mittels Qualitätsmodellen [32] in den Prozessen des Unternehmens nicht vorgesehen ist. Um allerdings den Katalog der Zusammenhänge des Leitfadens anzuwenden, wurden die Anforderungen im Rahmen dieser Arbeit trotz dessen konkretisiert und mit *Graphmasters* abgesprochen. Auch kommen nicht immer alle Evaluationsmöglichkeiten des Modells für einen Kontext infrage, da es äußere Beschränkungen gibt (siehe Abschnitt 6.2).

Als Einschränkung für die Allgemeingültigkeit des Leitfadens muss auch erwähnt werden, dass der Einsatz mit dieser Arbeit nur in einem agil arbeitenden Unternehmen im *Context* von mobilen Anwendungen untersucht wurde. Um den Nutzen der einzelnen Leitfadenteile allgemein beurteilen zu können, müssen folglich weitere Anwendungen des Leitfadens innerhalb von verschiedenen Prozessen und Kontexten erfolgen.

Eine weitere Einschränkung bei der Verallgemeinerung des Leitfadens ist, dass

keine direkte Evaluation des Leitfadens erfolgt ist. Diese ist nur indirekt durch die Anwendung des Leitfadens geschehen. Es können folglich Verständnisprobleme oder Herausforderungen bei der Nutzung in verschiedenen Prozessen bei der Entwicklung von Erklärungen auftreten.

Für die Verallgemeinerbarkeit des Leitfadens spricht, dass bei der Anwendung des Leitfadens mehrere Ergebnisse und Empfehlungen aus vorangegangenen Arbeiten reproduziert werden konnten. Beispielsweise konnte die Notwendigkeit von qualitativer und quantitativer Evaluation gezeigt werden, um interpretierbare Ergebnisse zu erlangen [37]. Auch konnte gezeigt werden, dass die Performanz der *End User* durch Erklärungen erhöht und gleichzeitig die *Satisfaction* sinken kann [62].

Final bieten die Ergebnisse einen allgemeinen Überblick für die Integration von Erklärungen in erklärbare Systeme. Somit ist es für Anwender des Leitfadens möglich, anhand von Einschränkungen des eigenen *Contexts* und äußeren Bedingungen die Teile des Leitfadens zu wählen, die ihnen bei der Integration von Erklärungen helfen können. Der Leitfaden bietet folglich eine gewisse Flexibilität.

7.3 Herausforderungen

Im Folgenden werden die Herausforderungen vorgestellt, die während der Forschung für diese Arbeit entstanden sind.

Eine der Hauptaufgaben dieser Arbeit ist das Zusammentragen der bestehenden Ergebnisse zur neuen NFR *Explainability* gewesen. Vor allem durch die sehr subjektiven und in verschiedenen Kontexten sehr unterschiedliche Wahrnehmung von Erklärungen war die Vereinheitlichung der Ergebnisse nicht trivial. Da vergangene Forschung vor allem einzelne Aspekte von Erklärungen ohne ein einheitliches Evaluationsschema analysiert hat, war auch die Aufstellung der Zusammenhänge zwischen den einzelnen Aspekten des in der vorliegenden Arbeit vorgestellten Leitfadens eine Herausforderung. Hier konnten vorwiegend existierende Modelle wie von Nunes und Jannach für Erklärungen für Empfehlungssysteme und die Definition für Erklärbarkeit von Chazette, Brunotte und Speith eine Orientierung geben [17, 5]. Aufgrund der Diversität von *Explainability* sind auch weniger allgemeine Resultate im Rahmen dieser Arbeit entstanden als zu Beginn der Literaturrecherche erwartet. Daher bietet der Leitfaden vordergründig einen Überblick über verschiedene Aspekte von Erklärungen sowie der Zusammenhänge.

Neben den Herausforderungen bei der Entwicklung des Leitfadens sind während dem Technologietransfer in die Praxis weitere Hürden zu überwinden gewesen. Der Leitfaden ist mit Zielen konzipiert, welche für die Anwendung in Qualitätsmodellen zur Ableitung von Anforderungen dienen. Die agile Arbeitsweise von *Graphmasters* ist allerdings nicht für eine konkrete Anforderungserhebung ausgelegt. Daher war diese Methode unbekannt. Folglich bedurfte die Festlegung der konkreten und überprüfbaren in Absprache mit *Graphmasters* einem besonderen Augenmerk. Insbesondere das Festlegen von Sollwerten für die entwickelten Metriken konnte nicht erfolgen. Hier wurde der Unterschied zwischen in der Wissenschaft entwickelten Modellen und Verwendung in der Wirtschaft deutlich gezeigt. Trotz dessen hatte der entwickelte Leitfaden eine unterstützende Wirkung für die Aktivitäten der Integration von Erklärungen (Anforderungserhebung, Umsetzung, Evaluation).

Auch konnten statistisch überprüfbare Qualitätsanforderungen formuliert werden (siehe Unterabschnitt 6.1.4). Für eine erneute Verwendung des Leitfadens bei *Graphmasters* muss allerdings eine bessere Integration in den agilen Prozess von *Graphmasters* erfolgen.

Kapitel 8

Fazit und Ausblick

8.1 Fazit

Ziel dieser Arbeit war es, ein Modell zur Unterstützung des Designs von Erklärungen in erklärbaren Systemen zu konzipieren und im Anschluss zu evaluieren. Ein Ergebnis dieser Arbeit ist folglich ein Modell, welches in einen Leitfaden integriert ist. Der Leitfaden enthält darüber hinaus einen Katalog über bestehende und verallgemeinerbare Zusammenhänge zwischen den äußeren Abhängigkeiten für Erklärungen, den Eigenschaften und Einflüssen auf die Softwarequalität. Abschließend werden im Leitfaden Heuristiken für das Design von Erklärungen zusammengefasst.

Der Leitfaden ist auf Basis einer Literaturrecherche entwickelt worden. Ziel dieser war es, externe Abhängigkeiten, Eigenschaften und Bewertungsmethoden von Erklärungen zu identifizieren, welche einen Einfluss auf ausgewählte Qualitätsaspekte haben. Diese Ergebnisse sind in dem im Leitfaden enthaltenen Modell zusammengefasst.

Ein weiteres Resultat dieser Arbeit ist die erfolgreiche Anwendung des vorgestellten Leitfadens in der Wirtschaft. Mithilfe des Leitfadens konnten im Rahmen eines Workshops zusammen mit der Firma *Graphmasters GmbH* aus Hannover Erklärungen in ein Navigationssystem integriert werden. Die Rohanforderungen, Umsetzungsideen und Ansätze zur Evaluation, welche das Ergebnis eines Workshops waren, wurden dazu mithilfe eines Qualitätsmodells in konkrete Anforderungen überführt. Final wurden diese in eine Produktivanwendung integriert und mit über 4 000 *End Usern* der Smartphone-Anwendung evaluiert. Zusammen mit einem anschließenden Quasi-Experiment mit vier Teilnehmern konnte geschlussfolgert werden, dass der Leitfaden bei der Integration von Erklärungen in ein bestehendes System zum einen die Entwicklung unterstützt hat. Zum anderen konnten positive Auswirkungen der Erklärungen im Rahmen der aufgestellten Ziele erreicht werden. Ferner konnte anhand der qualitativen Evaluation Verbesserungspotenzial der Erklärungen für weitere Iterationen aufgedeckt werden.

Zusammenfassend liefert diese Arbeit einen Leitfaden zur Integration von Erklärungen, welcher die Möglichkeit bietet, Erklärungen in ein System zu integrieren. Der Leitfaden hat während seiner Anwendung im Unternehmen bei allen wichtige Schritten der Integration von Erklärungen unterstützt (Anforderungserhebung, Umsetzung und Evaluation). Mit der Verwendung des Leitfadens konnten die

Erklärungen erfolgreich zum Erreichen von aufgestellten Qualitätszielen genutzt werden. Damit ist der Leitfaden ein erster Schritt bei der Entwicklung von Artefakten zur Vereinheitlichung der Entwicklung und Evaluation von Erklärungen in Wirtschaft und Wissenschaft [19, 80, 37].

Final erfüllt der in dieser Arbeit entwickelt Leitfaden folglich die Anforderung, eine Unterstützung bei der Gestaltung von Erklärungen in erklärbaren Systemen zu bieten.

8.2 Ausblick

Mit der Anwendung des entwickelten Leitfadens in der Wirtschaft konnte gezeigt werden, dass dieser das Ziel der Arbeit erfüllt. Die qualitative Evaluation ist zunächst lediglich mit vier Teilnehmern erfolgt, welche bereits verwendbare Rückmeldungen gegeben haben. Eine Ausweitung des Quasi-Experiments würde vermutlich zu weiteren und ggf. statistisch signifikanten Ergebnissen führen. Zudem sollte bei einer erneuten *Case Study* der Zeitraum verlängert werden, um mehr Studiengruppen mit gleichbleibender Nutzerzahl und zusätzliche Langzeitmetriken wie den Nutzerzuwachs zu ermöglichen.

Außerdem kann der einzelne Einsatz des Leitfadens nicht ausreichend belegen, dass der Leitfaden allgemeingültig anwendbar ist. Um eine erfolgreiche Verwendung in der Wirtschaft zu gewährleisten, sollte folglich im Anschluss an diese Arbeit ein strukturierter Technologietransfer, wie zum Beispiel von Gorschek et al. vorgestellt, durchgeführt werden [94]. Insbesondere könnte dieser Transfer weitere Verständnisprobleme des Leitfadens aufdecken. Je nach Anwendungsfall sollte außerdem in ein finales Artefakt mit dem Leitfaden eine Einführung in das Thema Erklärbarkeit integriert werden.

Zusätzlich sollte die Vollständigkeit des Modells für Erklärungen evaluiert und durch weitere unabhängige Arbeiten zu dem Thema bestätigt werden.

Ein Thema, welches im entwickelten Leitfaden wenig behandelt wird, sind mögliche negative Einflüsse, durch die Integration von Erklärungen. Es gibt einige Autoren, die unter anderem die Gefahr der Beeinträchtigung der *Usability* von Systemen durch Erklärungen diskutieren [5, 62, 19]. Eine Evaluation erfolgt aber in der Regel in Bezug auf positive Einflüsse von Erklärungen. Folglich sollten in zukünftigen Arbeiten negative Einflüsse näher betrachtet werden.

In diesem Zusammenhang fehlt außerdem ein Katalog über die genauen Einflüsse zwischen den verschiedenen Qualitätsaspekten, welche mit *Explainability* in Verbindung stehen. Chazette, Brunotte und Speith haben dafür bereits eine Grundlage mit einem Katalog der Aspekte, welche beeinflusst werden vorgestellt [5]. Auf Basis dessen sollte ein *Explainability Softgoal Interdependence Graph (SIG)* entwickelt werden. Dieser sollte über *Explainability* einen ähnlichen Überblick liefern wie bereits existierende SIGs über andere Qualitätsaspekte [vgl. 24, 1].

Folglich müssen in Zukunft weitere Artefakte entwickelt werden, welche die Möglichkeit bieten, die neue NFR *Explainability* in der Wirtschaft einfach anwendbar zu machen und die Betrachtung in der Wissenschaft zu vereinheitlichen [37]. Außerdem wird ein Prozess benötigt, der die entwickelten Methoden und Artefakte in bestehende Software-Entwicklungszyklen integriert [19, 74].

Anhang A

Literaturrecherche

Dieser Anhand enthält zusätzliche Materialien zur durchgeführten Literaturrecherche. Tabelle A.1 enthält die genaue Konfiguration der Suchen für die einzelnen Datenbanken und Abbildung A.1 die Anzahl der Ergebnisse pro Datenbank.

Eine Übersicht über die Kontexte, in welchen die resultierenden Arbeiten Erklärungen analysiert haben, ist in Abbildung A.1 zu finden.

Datenbank	Aktivierte Filter
ACM Digital Library	Content-Type: PDF
IEEE Xplore	Type: [Conferences, Journals] Subscribed Content Only
Science Direct	Article Type: [Research Articles, Book chapters] Subject Area: Computer Science
Springer Link	Discipline: Computer Science

Tabelle A.1: Zusätzliche Filter bei der Literaturrecherche in den Datenbanken

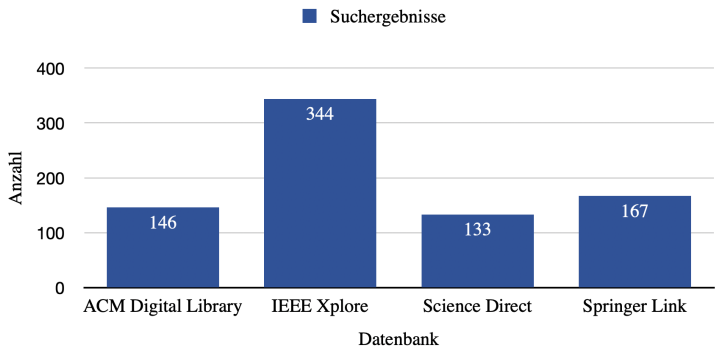


Abbildung A.1: Anzahl der Suchergebnisse in der Literaturrecherche pro Datenbank

Kontext	Quellen
Allgemein	[7] [2] [5] [45] [19] [78] [80]
Intelligente Systeme (z.B. XAI)	[69] [67] [37] [68] [20] [48] [49] [54] [65] [55] [30] [74] [75] [66] [76] [26] [77] [64] [57] [58] [23] [79] [25] [63]
Empfehlungssysteme	[43] [44] [22] [18] [46] [47] [35] [3] [51] [56] [17] [81]
Autonomes Fahren	[70] [52] [62] [61] [36] [73]
Mensch-Roboter-Interaktion	[71] [72] [53] [59] [60]
Spezifische Domänen	[50] [29]

Tabelle A.2: Kontexte von erklärbaren Systemen, die von in der Literaturrecherche betrachteten Arbeiten untersucht wurde

Anhang B

Integration von Erklärungen

In diesem Anhang sind zusätzliche Informationen zur Integration von Erklärungen in *NUNAV Navigation* enthalten.

Thema	Beschreibung	Anzahl
Feature Request	Im Review werden neue Funktionen gefordert.	18
Kollaboratives Routing	Das Review deutet auf ein fehlendes Verständnis, was der Grundgedanke des kollaborativen Routings ist, hin.	16
Schlechte Route	Das Review enthält eine Beschwerde über eine bestimmte Routenführung.	9
GPS-Verständnis	Das Review deutet auf ein fehlendes Verständnis für schlechten GPS-Empfang hin.	3
Offline-Modus-Verständnis	Das Review deutet auf ein fehlendes Verständnis der Offline-Karten-Funktion hin.	3
Fehlende Information	Das Review enthält eine Forderung nach mehr Informationen, die in NUNAV angezeigt werden sollen.	3
Schlechte Suchergebnisse	Im Review wird die Qualität der Suchergebnisse bemängelt.	3
Verständnis Routenfarbe	Im Review werden Nachfragen zur Einfärbung der Route gestellt.	2
Falsche Kartendaten	Im Review werden falsche Kartendaten bemängelt.	1

Tabelle B.1: Anzahl der Reviews im Google Play Store und Apple App Store mit mehrfach vorkommenden Themen

Workshop zur Integration von Erklärungen

Workshop: Erklärungen in NUNAV - Wie können Erklärungen das Nutzererlebnis verbessern?

📋 Übersicht

Workshop Datum	**06 Jul 2021**
Titel	Erklärungen in NUNAV – Wie können Erklärungen das Nutzererlebnis verbessern?
Ziele	1. Verstehen von Erklärbarkeit als Nicht-Funktionale Anforderung (NFR): Welche Probleme können durch Erklärbarkeit gelöst werden? 2. Klarstellen von Verständnis-Problemen in NUNAV und herausarbeiten von Zielen, die erreicht werden sollen. 3. Lösungsansätze du den Problemen sammeln, die durch Erklärungen gelöst werden können. 4. Evaluationsmethoden zusammenstellen, durch welche die Ideen überprüft werden können.
Moderator	Florian Herzog
Teilnehmer	[redacted] – Lead Mobile Development [redacted] – Mobile Development Team [redacted] – Mobile Development Team [redacted] – Product Manager [redacted] – Web Development Team [redacted] – Web Development Team [redacted] – Solution Experts Team
Dauer	ca. 2,5h

🕛 Ablauf

Nr.	Titel	Verantwortlich	Beschreibung
1	Einleitung Erklärbarkeit	Florian	• Einführung in Erklärbarkeit als NFR • Motivation, warum Erklärungen in mobilen Anwendungen nützlich sein können • Vorstellung des ausgearbeiteten Modells • Beispiele für Erklärungen geben
2	Vorstellung von bekannten Problemen	Florian	• Bekannte Fragen und Unverständlichkeiten, die bei Nutzern aufgetreten sind, werden vorgestellt • Die Daten wurden dabei aus folgenden Quellen extrahiert: 1. Nutzerfeedback (Play Store, App Store, In-App) 2. Supportanfragen 3. Aufrufe von Support-Artikeln
	Weitere Probleme sammeln und ordnen	Alle	• Die bestehenden Probleme in Kategorien einordnen • Die Probleme priorisieren
3	Nutzer festlegen	Alle	• Verschiedene Nutzerprofile sammeln
4	Festlegen der Ziele und Brainstorming für die Umsetzung	Alle	• Klarstellen, was erreicht werden soll • Von der Businessebene bis zur Erklärung • Die einzelnen Aspekte der Erklärungen festlegen
5	Metriken zur Überprüfung suchen	Alle	• Ideen für Metriken sammeln • Umsetzbarkeit der Metriken überprüfen
6	Umsetzung festmachen	Alle	• Festlegen, welche Ziele verfolgt werden sollten • Festlegen, welche Erklärungsideen vertieft werden sollen • Metriken festlegen

🚀 Ergebnisse

1. Einleitung Erklärbarkeit

Rückfragen

- Wo ist die Grenze zwischen gutem UI / UX und Erklärbarkeit?
    - Frage: “Ist das Weglassen von Buttons, um das Interface zu vereinfachen, Erklärbarkeit?”
    - Antwort: “Für Erklärbarkeit muss das System mit dem Nutzer kommunizieren.“

  Hätte besser in der Präsentation kommuniziert werden müssen

- Einzelne Rückfragen zu bestimmten Punkten des Modells
    - Besonders auffällig war, dass vor allem die verschiedenen Informationstypen nicht klar genug definiert sind

2. Häufigste Probleme und Nutzerfragen

1. Navigation:
    a. Kollaboratives Routing
        i. Was ist / wie funktioniert kollaboratives Routing?
        ii. Warum brauche ich eine ständige Internetverbindung?
    b. Vertrauen in das System
        i. Woher kommen die Verkehrs- und Sperrungs-Daten?
        ii. Was passiert mit meinen Daten?
        iii. Routen
            - Warum wird diese Route / dieser Umweg genommen?
            - Warum sind die Routen bei NUNAV länger?
        iv. Nutzer kann die Güte der Route nicht bewerten
    c. Sonstige
        i. Warum bekomme ich keine Route?
2. Funktionen:
    a. Was bedeutet die Routenfarbe?
    b. Wie kann ich Favoriten anlegen?
    c. Wo kann ich mein Ziel in NUNAV eingeben?
    d. Wie kann ich eine Sperrung melden?
    e. Warum ist die Position ungenau?

3. Nutzer festlegen

- “Poweruser”: Verwendet täglich ein Navigationssystem
- “Event-User“: Nutzt NUNAV erstmalig als Navigationssystem im Kontext von Venue-Navigation
- ~~“Courier“: Nutzt Courier täglich zum Pakete ausliefern~~ (Außerhalb der Arbeit)

4. Festlegen der Ziele und Brainstorming für die Umsetzung

Ziele

Hauptziel:

1. Möglichst viele Nutzer von NUNAV überzeugen.

Verständnisprobleme beim Nutzer lösen:

1. Der Nutzer soll mit seiner Route zufrieden sein
2. Der Nutzer soll sich an die vorgeschlagene Route halten

Lösungsansätze

1. Mehr Kontextinformationen geben
2. Der Nutzer soll verstehen, was kollaboratives Routing ist.

Ideen für Erklärungen

Idee	Argumentation	Aspekt	Aufwand
Alternativrouten anzeigen (nicht auswählbar)	+ Wenn dem Nutzer die Route ohne unseren Algorithmus (keine Verteilung, keine Reservierung, kein Live-Verkehr), versteht er, warum diese schlechter ist (Verkehrsfluss anzeigen) - Tiefer Systemeingriff, muss länger als im Rahmen der Masterarbeit diskutiert werden und im Rahmen unserer Vision geprüft werden	• 1.a.i + ii • 1.b.iii + iv	HOCH

Andere Nutzer auf der Karte anzeigen	+ Gibt gutes Verständnis für den Schwarmalgorithmus - Gutes Design sehr aufwendig - Datenschutztechnisch aktuell nicht umsetzbar - Hoher Datenverbrauch	• 1.a.i + ii	MITTEL
Detailliertere Fehlermeldungen	+ Der Nutzer versteht besser, warum er keine Route zu einem bestimmten Ziel bekommt - Sehr spezifisch auf eine Funktion zugeschnitten - Daten liegen aktuell im Backend nicht (immer) vor	• 1.c.i	MITTEL
Funktionen genau erklären	+ Einfach einzubauen + Support-Artikel existieren bereits - Zu wenig Anfragen zur Auswertung - Nicht Kernproblem	• 2a-e	LEICHT
Banner mit Erklärungen zu FAQs anzeigen	+ Deckt viele Nutzerfragen ab + Einfach zu integrieren - Kann störend wirken	• 1.a.i + ii • 1.b.i + ii	LEICHT
Sprachansagen während dem Routing hinzufügen	+ Vielseitig einsetzbar + Einfach zu erweitern - Mögliche Ablenkung während der Navigation - Sehr vielfältig und nicht genau prüfbar	• 1.b.iii + iv	MITTEL
Variable Sprachansagen bei Stau	+ Klare Begründung / Vergewisserung für den Nutzer - Nicht immer richtig erkennbar	• 1.b.iii + iv	MITTEL
Audioansagen bei schlechtem GPS integrieren	+ Ermöglicht dem Nutzer einzuschätzen, inwiefern das System gerade fehlerhaft ist - Definition von “Schlechtem GPS“ schwierig	• 2.e	MITTEL
Anzeigen der aktuellen Nutzerzahl im System	+ Vermutlich leichte Möglichkeit den kollaborativen Ansatz zu vermitteln - Kann je nach Designumsetzung unruhig wirken	• 1.a.i	LEICHT
Nach einem Absturz aufklären	+ Gibt dem Nutzer zusätzliche Kontrolle + Zusätzliches Feedback möglich - Führt ggf. zu einem erneuten ins Gedächtnis rufen des Fehlers	• -	MITTEL
Am Ende der Route einen Vergleich zwischen Vorhersage und gefahrener Route geben	+/- Gibt dem Nutzer einen tiefen Einblick in die Qualität des Systems - Das Aufzeichnen des genauen Fahrtweges und die Bewertung im Nachgang ist schwierig umsetzbar	• 1.b.iii + iv	HOCH
Mehr Kontextinformationen auf der Karte anzeigen	+ Der Nutzer kann sich die Route an weiteren Kontextinformationen, wie z. B. vielen Ampeln die Route selbst verständlicher erklären - Es kann zu einer hohen Unübersichtlichkeit auf der Karte kommen - Auswahl der genauen Daten unklar	• 1.b.iii + iv	MITTEL

Weiterer Klärungsbedarf

- Welchen Ton sollen unsere Erklärungen treffen?
  - Analytisch, Faktisch, Albern, Ernst
- ENTSCHEIDUNG Faktisch

5. Metriken

1. Messen, wie stark sich der Nutzer an die vorgegebene Route hält (Wie oft hat der Nutzer eine neue Route forciert)
2. Messen, wie zufrieden der Nutzer mit der Route war (Idee: ggf. Frage am Ende der Route ändern)
3. Wie bewertet der Nutzer bei Anzeige der Support-Artikel deren Nützlichkeit
4. Wie oft verwendet der Nutzer die App

6. Umsetzung festmachen

Erklärung	Schritte
Audioansagen zu Beginn der Route integrieren	Klären, welche Daten genutzt werden können (Backend) Bedarf definieren Strings definieren
Audioansagen am Ende der Route integrieren	• Klären, welche Daten genutzt werden können (Backend) • Bedarf definieren • Strings definieren
Audioansagen bei schlechtem GPS integrieren	• Umsetzung prüfen • Strings definieren
Audioansagen bei Routenänderung integrieren	• Klären, welche Daten genutzt werden können (Backend) • Bedarf definieren • Strings definieren
Anzeige von Bannern mit Links zu den FAQs	• Klären, welche FAQs umgesetzt werden • Stelle, an der diese angezeigt werden, klären • Strings definieren • Support-Artikel aktualisieren • UI designen
Anzeigen der aktuellen Nutzerzahl im System	• Klären, welche Nutzerzahl angezeigt werden soll (bestimmter Umkreis oder alle) • UI designen
Metriken im System zusammenführen	• Metriken klar definieren • Aktuelles Feedback anzeigen • Übertragungsformat definieren

Weitere Aufgaben

Aufgabe	
Klarstellung der Definition von Erklärbarkeit	• Recherche in bestehenden Aufzeichnungen für neue Titel / Definitionen / Aufteilungen durchführen • Ergebnisse in das Modell einpflegen
Überarbeitung der Informationstypen im Modell	• Recherche in bestehenden Aufzeichnungen für neue Titel / Definitionen / Aufteilungen durchführen • Ergebnisse in das Modell einpflegen

Hife-Center-Artikel: Collaborative Routing

https://support.graphmasters.net/hc/articles/360010648960-Was-ist-Collaborative-Routing-,
besucht: 14.10.21

Wir geben jedem Nutzer eine individuelle Route, um an sein Ziel zu kommen. Diese Route steht immer in Abhängigkeit zu allen anderen Nutzern und sorgt dafür, dass alle Fahrzeuge bestmöglich auf die bestehende Infrastruktur verteilt werden. Durch diese Schwarmintelligenz sind Routingmodelle möglich, die weit über das herkömmliche „Kürzester Weg"-Modell hinausgehen und so Staus aktiv vermeiden.

Man nennt diese Art des Routings „Collaborative Routing".

Hife-Center-Artikel: Routeneinflüsse

https://support.graphmasters.net/hc/de/articles/4404558277266-Wie-ist-meine-NUNAV-Route-entstanden-,
besucht: 14.10.21

Jede berechnete Route entstammt dem neuartigen Ansatz des Collaborative Routings. Dabei werden eine Vielzahl von Variablen berücksichtigt:

- Die Routen von NUNAV-Nutzern in Ihrer Nähe
- Sperrungen
- Baustellen
- Aktuelle Verkehrsverzögerungen
- und viele mehr...

Anders als bei herkömmlichen Navigationssystemen ist eine **NUNAV-Route** keinesfalls statisch. Aufgrund der dynamischen Veränderung des Verkehrs berechnen wir Ihnen mehrfach in der Minute Ihre aktuell schnellste individuelle Route unter Berücksichtigung der aktuellen Verkehrslage.

Das führt dazu, dass **NUNAV-Route** intelligent und stauvermeidend auf das vorhandene Straßennetz verteilt werden.

Bei alternativen Navigationssystemen wird meistens egoistisch gehandelt, es oft geht darum, deren Nutzer auf dem kürzesten Wege zu seinem Ziel zu bringen. Dieser Ansatz ist keineswegs stauvermeidend, sondern verursacht oft das Gegenteil.

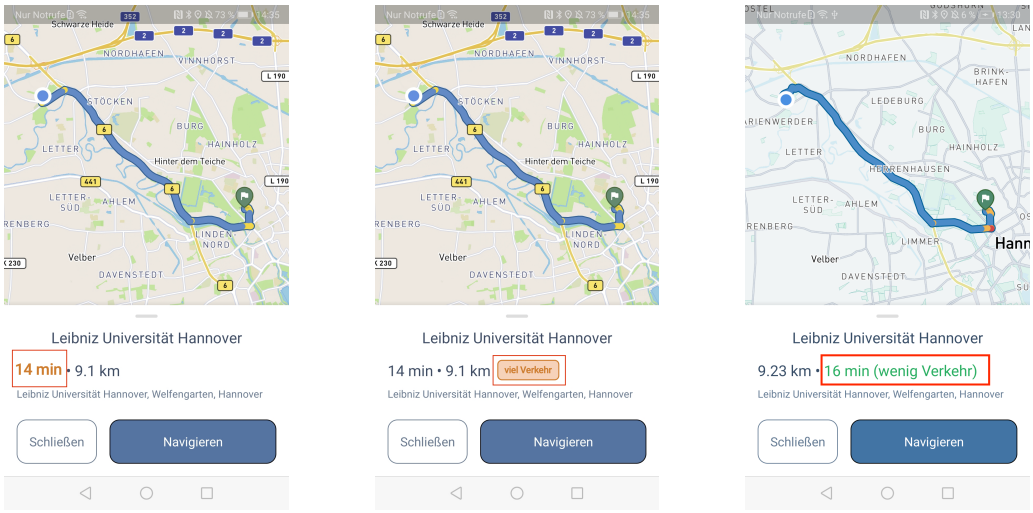
Bei NUNAV hingegen, sind Sie Teil einer Gemeinschaft, welche alle Nutzer schnellstmöglich und stauvermeidend an Ihr Ziel bringt. Die wichtigsten Vorteile von **NUNAV-Route** lauten daher wie folgt:

- Stauvermeidend durch den Ansatz des Collaborative Routings
- CO2 Einsparung durch weniger Stau und Fahrtzeit
- Gemeinschaftliches Handeln führt zu einer Verbesserung der generellen Verkehrslage

Je mehr Nutzer, sich mit **NUNAV-Route** zu Ihrem Ziel navigieren lassen, desto größer ist der Effekt, den wir gemeinsam erzielen können. Deswegen würden wir uns freuen, wenn Sie uns an Ihre Freunde und Familien weiterempfehlen.

Wir hoffen, dass wir Ihnen mit diesem Artikel die Entstehung Ihrer individuellen **NUNAV-Route** etwas näher bringen konnten.

Prototypen der Erklärung zum aktuellen Verkehrsaufkommen

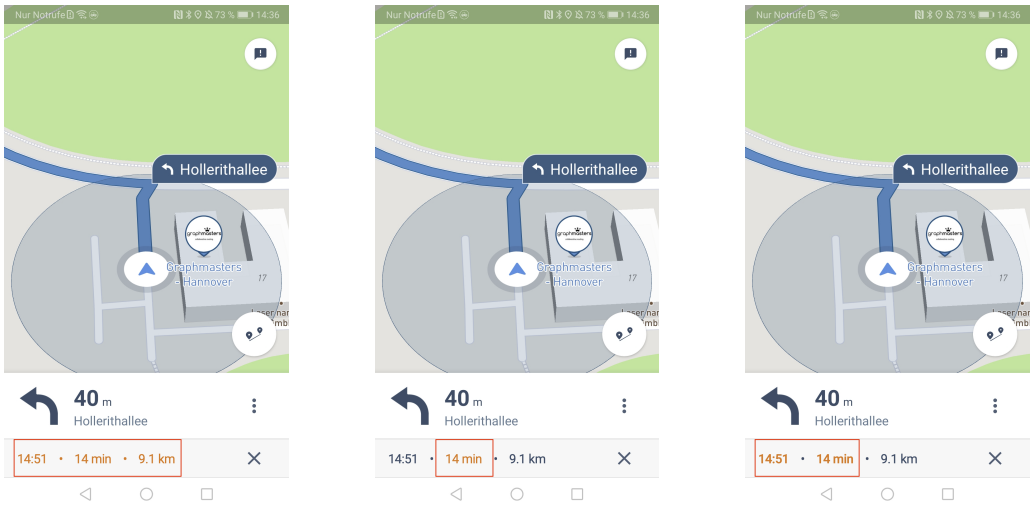


(a) 1. Prototyp zur Darstellung des aktuellen Verkehrsaufkommens

(b) 2. Prototyp zur Darstellung des aktuellen Verkehrsaufkommens

(c) Finales Design zur Darstellung des aktuellen Verkehrsaufkommens

Abbildung B.1: Prototyp und finale Designs für die Erklärung zum aktuellen Verkehrsaufkommen in der Routenvorschau



(a) 1. Prototyp zur Darstellung des aktuellen Verkehrsaufkommens

(b) 2. Prototyp zur Darstellung des aktuellen Verkehrsaufkommens

(c) Finales Design zur Darstellung des aktuellen Verkehrsaufkommens

Abbildung B.2: Prototyp und finale Designs für die Erklärung zum aktuellen Verkehrsaufkommen während der Navigation

Anhang C

Evaluation der Erklärungen

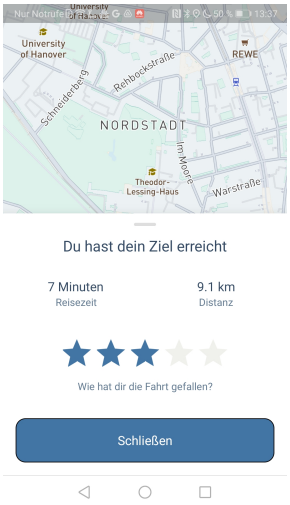


Abbildung C.1: Bildschirmfoto des Bewertungsdialogs für die Zufriedenheit mit der Navigation

Fragebogen des Quasi-Experiments

Einwilligungs- und Datenschutzerklärung

zur Teilnahme am Software-Engineering-Experiment:

„Konzeption und Evaluation eines Modells zur Unterstützung des Designs von Erklärungen in erklärbaren Systemen“

Name/Vorname des/der Teilnehmers/in: _____

1. Ich bin über die Art, Durchführung und etwaige Risiken des Experiments aufgeklärt worden. Alle meine Fragen zum vorgesehenen Experiment wurden zu meiner Zufriedenheit beantwortet.
2. Ich versichere hiermit, dass ich sämtliche Fragen nach bestem Wissen beantworte und dass ich mich an die Anweisungen im Rahmen dieses Experiments halten werde.
3. Die Teilnahme an diesem Experiment ist freiwillig. Ich kann jederzeit ohne Angabe von Gründen und ohne Nachteile meine Teilnahme widerrufen.
4. Ich erkläre mich bereit, im Rahmen des Experiments an einem Interview mit dem/der Durchführenden des Experiments teilzunehmen.
5. Mir ist bewusst, dass mit dem Ausfüllen der Einwilligungs- und Datenschutzerklärung kein verbindliches Recht auf Teilnahme am Experiment verbunden ist.
6. Alle im Rahmen des Experiments erhobenen Daten werden ausschließlich anonym, d.h. ohne Nennung des Namens oder anderer personenbezogener Merkmale, behandelt, verarbeitet, weitergegeben oder veröffentlicht. Nach Abschluss des Experiments werden alle personenbezogenen Daten gelöscht. Die Daten werden ausschließlich zu wissenschaftlichen Zwecken verwendet.
7. Das Interview wird aufgezeichnet. Die Aufzeichnung wird nur für Forschungszwecke verwendet und nach der Evaluation von den Servern gelöscht. Die Aufzeichnung wird nicht veröffentlicht.
8. Die ordnungsgemäße Durchführung des Experiments, insbesondere die ordnungsgemäße Erhebung der Daten sowie deren Zuordnung zu bestimmten Studienteilnehmern, wird von Mitarbeitern/innen überprüft, die durch den/die Leiter/in des Experiments dazu autorisiert wurden. Nur diese Mitarbeiter/innen haben direkte Einsicht in die personenbezogenen Daten und haben sich zum Stillschweigen verpflichtet.

Ich stimme der oben beschriebenen Erfassung und Behandlung meiner personenbezogenen Daten und Einsichtnahme in diese Daten zu.

……..….	……...……..
Ort, Datum und Unterschrift des/der Teilnehmers/in	Ort, Datum und Unterschrift des/der Durchführenden

Durchführende/r und Ansprechpartner: Florian Herzog, Fachgebiet Software Engineering, Welfengarten 1, 30167 Hannover; 0511-762-4793; florian.herzog@graphmasters.net

Leiter/in des Experiments: Larissa Chazette, Fachgebiet Software Engineering, Welfengarten 1, 30167 Hannover; 0511-762-4793; larissa.chazette@inf.uni-hannover.de

1. Allgemeine Angaben

1.1 Geschlecht

[ ] Weiblich
[ ] Männlich
[ ] Divers

1.3 Alter

_____ Jahre

1.2 Nutzung von Navigationssystemen

Bitte bewerten Sie die folgenden Aussagen:

	Volle Ablehnung				Volle Zustimmung
Ich bin mit Navigationsanwendungen wie z.B. Google Maps, Apple Maps, etc. Vertraut.	◯	◯	◯	◯	◯
	< 1 mal	1 mal	2 mal	3 mal	> 3 mal
Ich nutze Navigationsanwendungen durchschnittlich pro Woche	◯	◯	◯	◯	◯

1.3 NUNAV Navigation

In diesem Experiment sollen Teile der App NUNAV Navigation evaluiert werden. NUNAV ist eine mit Apple oder Google Maps vergleichbare Navigationsanwendung.

Allerdings verwendet NUNAV einen sogenannten kollaborativen Routing-Algorithmus, der die NUNAV-Nutzer besser auf die vorhandene Infrastruktur verteilt als bei herkömmlichen Navigationssystemen.

Bitte bewerten Sie die folgenden Aussage:

	Gar nicht	1-2	3-5	5-10	> 10
Ich habe NUNAV Navigation schon …. Mal zur Navigation verwendet.	◯	◯	◯	◯	◯

Im folgenden werden Ihnen verschiedene Erklärungen präsentiert, die in NUNAV Navigation integriert wurden.

Sie sollen zu jeder Erklärung zunächst in eines sogenannten Think-Aloud-Parts erzählen, was sie über diese Erklärung denken und damit interagieren.

Das heißt Sie sollen alle Gedanken wie beispielsweise aufkommende Fragen laut aussprechen.

Im Anschluss an jede Erklärung erhalten Sie einen kurzen Fragebogen, der sich auf die jeweils zuvor gezeigte Erklärung bezieht.

Nachdem Sie alle Erklärungen gesehen haben, werden Ihnen in einem offenen Gespräch Rückfragen zu ihren Antworten oder Ausführungen gestellt.

2 Erklärungen

2.1 Kollaboratives Routing

Ziel der Erklärung: Sie sollen verstehen, was kollaboratives Routing ist und wie es funktioniert.

Bitte bewerten Sie die folgenden Aussagen:

	Volle Ablehnung				Volle Zustimmung
Die Erklärung stellt mich mit meinem Verständnis über kollaboratives Routing zufrieden.	◯	◯	◯	◯	◯
Die Informationen der Erklärung sind ausreichend für die Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist überzeugend.	◯	◯	◯	◯	◯
Die Erklärung weckt Interesse.	◯	◯	◯	◯	◯
Die Erklärung ist einfach zu verstehen.	◯	◯	◯	◯	◯
Die Erklärung hilft bei der Verwendung von NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist nützlich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist hinreichend detailliert.	◯	◯	◯	◯	◯
Die Erklärung ist zu lang.	◯	◯	◯	◯	◯

	Volle Ablehnung				Volle Zustimmung
Ich habe diese Erklärung benötigt.	◯	◯	◯	◯	◯
Ich würde mir diese Erklärung mehrfach ansehen.	◯	◯	◯	◯	◯
Die Erklärung stört mich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Ich möchte diese Erklärung ausblenden können.	◯	◯	◯	◯	◯

Fehlende Informationen:

Verbesserungsvorschläge:

Sonstiges:

2.2 Meine Route

Ziel der Erklärung: Sie sollen verstehen, welche äußeren Einflüsse auf die Routenberechnung wirken.

Bitte bewerten Sie die folgenden Aussagen:

	Volle Ablehnung				Volle Zustimmung
Die Erklärung stellt mich mit meinem Verständnis über die Einflüsse auf das Routing zufrieden.	◯	◯	◯	◯	◯
Die Informationen der Erklärung sind ausreichend für die Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist überzeugend.	◯	◯	◯	◯	◯
Die Erklärung weckt Interesse.	◯	◯	◯	◯	◯
Die Erklärung ist einfach zu verstehen.	◯	◯	◯	◯	◯
Die Erklärung hilft bei der Verwendung von NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist nützlich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist hinreichend detailliert.	◯	◯	◯	◯	◯
Die Erklärung ist zu lang.	◯	◯	◯	◯	◯

	Volle Ablehnung				Volle Zustimmung
Ich habe diese Erklärung benötigt.	◯	◯	◯	◯	◯
Ich würde mir diese Erklärung mehrfach ansehen.	◯	◯	◯	◯	◯
Die Erklärung stört mich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Ich möchte diese Erklärung ausblenden können.	◯	◯	◯	◯	◯

Fehlende Informationen:

Verbesserungsvorschläge:

Sonstiges:

2.3 Verkehr auf der Route

Ziel der Erklärung: Sie sollen sehen können, welchen Einfluss der Verkehr bei der Berechnung der Route spielt.

Bitte bewerten Sie die folgenden Aussagen:

	Volle Ablehnung				Volle Zustimmung
Die Erklärung stellt mich mit meinem Verständnis über den Einfluss der Verkehrs auf das Routing zufrieden.	◯	◯	◯	◯	◯
Die Informationen der Erklärung sind ausreichend für die Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist überzeugend.	◯	◯	◯	◯	◯
Die Erklärung weckt Interesse.	◯	◯	◯	◯	◯
Die Erklärung ist einfach zu verstehen.	◯	◯	◯	◯	◯
Die Erklärung hilft bei der Verwendung von NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist nützlich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist hinreichend detailliert.	◯	◯	◯	◯	◯
Die Erklärung ist zu lang.	◯	◯	◯	◯	◯

	Volle Ablehnung				Volle Zustimmung
Ich habe diese Erklärung benötigt.	◯	◯	◯	◯	◯
Ich würde mir diese Erklärung mehrfach ansehen.	◯	◯	◯	◯	◯
Die Erklärung stört mich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Ich möchte diese Erklärung ausblenden können.	◯	◯	◯	◯	◯

Fehlende Informationen:

Verbesserungsvorschläge:

Sonstiges:

2.4 Ungenaue Positionierung während der Navigation

Ziel der Erklärung: Sie sollen verstehen, wann die Navigation mit NUNAV unzuverlässig ist.

Bitte bewerten Sie die folgenden Aussagen:

	Volle Ablehnung				Volle Zustimmung
Die Erklärung stellt mich mit meinem Verständnis über die Verlässlichkeit der Navigation zufrieden.	◯	◯	◯	◯	◯
Die Informationen der Erklärung sind ausreichend für die Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist überzeugend.	◯	◯	◯	◯	◯
Die Erklärung weckt Interesse.	◯	◯	◯	◯	◯
Die Erklärung ist einfach zu verstehen.	◯	◯	◯	◯	◯
Die Erklärung hilft bei der Verwendung von NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist nützlich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Die Erklärung ist hinreichend detailliert.	◯	◯	◯	◯	◯
Die Erklärung ist zu lang.	◯	◯	◯	◯	◯

	Volle Ablehnung				Volle Zustimmung
Ich habe diese Erklärung benötigt.	◯	◯	◯	◯	◯
Ich würde mir diese Erklärung mehrfach ansehen.	◯	◯	◯	◯	◯
Die Erklärung stört mich bei der Navigation mit NUNAV.	◯	◯	◯	◯	◯
Ich möchte diese Erklärung ausblenden können.	◯	◯	◯	◯	◯

Fehlende Informationen:

Verbesserungsvorschläge:

Sonstiges:

Literatur

[1] Rainara Maia Carvalho, Rossana MC Andrade und Káthia M Oliveira. „How developers believe Invisibility impacts NFRs related to User Interaction“. In: *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE. 2020, S. 102–112. DOI: 10.1109/RE48521.2020.00022.

[2] Larissa Chazette und Kurt Schneider. „Explainability as a non-functional requirement: challenges and recommendations“. In: *Requirements Engineering* 25.4 (2020), S. 493–514. DOI: 10.1007/s00766-020-00333-1.

[3] Nava Tintarev und Judith Masthoff. „Explaining recommendations: Design and evaluation“. In: *Recommender systems handbook*. Springer, 2015, S. 353–382. DOI: 10.1007/978-1-4899-7637-6_10.

[4] Mario A Cypko et al. „A guide for constructing bayesian network graphs of cancer treatment decisions“. In: *MEDINFO 2017: Precision Healthcare through Informatics*. IOS Press, 2017, S. 1355–1355. DOI: 10.3233/978-1-61499-830-3-1355.

[5] Larissa Chazette, Wasja Brunotte und Timo Speith. *A Knowledge Catalogue for Explainability: Definitions, Impacts, and Dimensions*. 2021.

[6] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *ISO/IEC 25010: Systems and software engineering-systems and Software Quality Requirements and Evaluation (SQuaRE)*. 2011.

[7] Larissa Chazette, Oliver Karras und Kurt Schneider. „Do End-Users Want Explanations? Analyzing the Role of Explainability as an Emerging Aspect of Non-Functional Requirements“. In: 2019, S. 11. DOI: 10.1109/RE.2019.00032.

[8] Europäisches Parlament und Rat der euopäischen Union. *Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)*. 2016. URL: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679&qid=1629269806673&from=EN#d1e40-1-1 (besucht am 18. 08. 2021).

[9] Consumer Financial Protection Bureau. *12 CFR Part 1002 - Equal Credit Opportunity Act (Regulation B)*. 2018. URL: https://www.consumerfinance.gov/rules-policy/regulations/1002/9/ (besucht am 18. 08. 2021).

[10] Adam Mosseri. *Shedding More Light on How Instagram Works*. 2021. URL: https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works (besucht am 18. 08. 2021).

[11] TikTok Technology Limited. *How TikTok recommends videos #ForYou*. 2021. URL: https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you/ (besucht am 18. 08. 2021).

[12] Ruth MJ Byrne. „The construction of explanations“. In: *AI and Cognitive Science'90*. Springer, 1991, S. 337–351. DOI: 10.1007/978-1-4471-3542-5_21.

[13] Alison Cawsey. „Generating Interactive Explanations“. In: *AAAI*. Citeseer. 1991, S. 86–91.

[14] Leilani H. Gilpin et al. „Explaining Explanations: An Overview of Interpretability of Machine Learning“. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. 2018, S. 80–89. DOI: 10.1109/DSAA.2018.00018.

[15] Ruth C. Fong und Andrea Vedaldi. „Interpretable Explanations of Black Boxes by Meaningful Perturbation“. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504. 2017, S. 3449–3457. DOI: 10.1109/ICCV.2017.371.

[16] Wojciech Samek und Klaus-Robert Müller. „Towards Explainable Artificial Intelligence“. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Lecture Notes in Computer Science. Springer International Publishing, 2019, S. 5–22. ISBN: 978-3-030-28954-6. DOI: 10.1007/978-3-030-28954-6_1. (Besucht am 24. 05. 2021).

[17] Ingrid Nunes und Dietmar Jannach. „A systematic review and taxonomy of explanations in decision support and recommender systems“. In: *User Modeling and User-Adapted Interaction* 27.3 (2017), S. 393–444. ISSN: 1573-1391. DOI: 10.1007/s11257-017-9195-0. (Besucht am 24. 05. 2021).

[18] Pigi Kouki et al. „User Preferences for Hybrid Explanations“. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. RecSys '17. event-place: Como, Italy. New York, NY, USA: Association for Computing Machinery, 2017, S. 84–88. ISBN: 978-1-4503-4652-8. DOI: 10.1145/3109859.3109915.

[19] Maximilian A. Köhl et al. „Explainability as a Non-Functional Requirement“. In: *2019 IEEE 27th International Requirements Engineering Conference (RE)*. ISSN: 2332-6441. 2019, S. 363–368. DOI: 10.1109/RE.2019.00046.

[20] Andrea Brennen. „What Do People Really Want When They Say They Want Explainable AI? We Asked 60 Stakeholders.“ In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. event-place: Honolulu, HI, USA. New York, NY, USA: Association for Computing Machinery, 2020, S. 1–7. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3383047.

[21] Zhongpin Wang. „Integration and Evaluation of Explanations in the Context of a Navigation App“. Bachelor. Leibniz University Hanover, 2020.

[22] Krisztian Balog und Filip Radlinski. „Measuring Recommendation Explanation Quality: The Conflicting Goals of Explanations“. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '20. event-place: Virtual Event, China. New York, NY, USA: Association for Computing Machinery, 2020, S. 329–338. ISBN: 978-1-4503-8016-4. DOI: 10.1145/3397271.3401032.

[23] Francesco Sovrano, Fabio Vitali und Monica Palmirani. „Modelling GDPR-Compliant Explanations for Trustworthy AI“. In: *Electronic Government and the Information Systems Perspective*. Hrsg. von Andrea Kő et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 219–233. ISBN: 978-3-030-58957-8. DOI: 10.1007/978-3-030-58957-8_16.

[24] Julio Cesar Sampaio do Prado Leite und Claudia Cappelli. „Software transparency“. In: *Business & Information Systems Engineering* 2.3 (2010), S. 127–139. DOI: 10.1007/s12599-010-0102-z.

[25] Finale Doshi-Velez und Been Kim. *Towards a rigorous science of interpretable machine learning*. 2017.

[26] Robert Thomson und Jordan Richard Schoenherr. „Knowledge-to-Information Translation Training (KITT): An Adaptive Approach to Explainable Artificial Intelligence“. In: *Adaptive Instructional Systems*. Hrsg. von Robert A. Sottilare und Jessica Schwarz. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 187–204. ISBN: 978-3-030-50788-6. DOI: 10.1007/978-3-030-50788-6_14.

[27] Michelene TH Chi. „Three types of conceptual change: Belief revision, mental model transformation, and categorical shift“. In: *International handbook of research on conceptual change*. Routledge, 2009, S. 89–110. ISBN: 9780203874813.

[28] Donald A Norman. *The psychology of everyday things*. Basic books, 1988. DOI: 10.1016/0004-3702(89)90083-0.

[29] Zahra Zahedi et al. „Towards Understanding User Preferences for Explanation Types in Model Reconciliation“. In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Daegu, Korea (South): IEEE, 2019, S. 648–649. ISBN: 978-1-5386-8555-6. DOI: 10.1109/HRI.2019.8673097. (Besucht am 24. 05. 2021).

[30] Avi Rosenfeld und Ariella Richardson. „Explainability in human–agent systems“. In: *Autonomous Agents and Multi-Agent Systems* 33.6 (2019), S. 673–705. ISSN: 1573-7454. DOI: 10.1007/s10458-019-09408-y. (Besucht am 24. 05. 2021).

[31] Lawrence Chung und Julio Cesar Sampaio do Prado Leite. „On non-functional requirements in software engineering“. In: *Conceptual modeling: Foundations and applications*. Springer, 2009, S. 363–379. DOI: 10.1007/978-3-642-02463-4_19.

[32] Kurt Schneider. *Abenteuer Softwarequalität: Grundlagen und Verfahren für Qualitätssicherung und Qualitätsmanagement*. dpunkt. verlag, 2012. ISBN: 3898647846.

[33] IEEE Computer Society. *IEEE 1061-1992 - IEEE Standard for a Software Quality Metrics Methodology*. 1992. URL: https://standards.ieee.org/standard/1061-1992.html (besucht am 23. 09. 2021).

[34] Claes Wohlin et al. *Experimentation in software engineering*. Springer Science & Business Media, 2012. ISBN: 3642290434.

[35] Johannes Kunkel et al. „Let Me Explain: Impact of Personal and Impersonal Explanations on Trust in Recommender Systems“. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. event-place: Glasgow, Scotland Uk. New York, NY, USA: Association for Computing Machinery, 2019, S. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300717.

[36] Gesa Wiegand et al. „I drive-you trust: Explaining driving behavior of autonomous cars“. In: *Extended abstracts of the 2019 chi conference on human factors in computing systems*. 2019, S. 1–6. DOI: 10.1145/3290607.3312817.

[37] Kacper Sokol und Peter Flach. „Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches“. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT* '20. event-place: Barcelona, Spain. New York, NY, USA: Association for Computing Machinery, 2020, S. 56–67. ISBN: 978-1-4503-6936-7. DOI: 10.1145/3351095.3372870.

[38] Christian J Mahoney et al. „A Framework for Explainable Text Classification in Legal Document Review“. In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. 2019, S. 1858–1867. DOI: 10.1109/BigData47090.2019.9005659.

[39] Upol Ehsan et al. „Operationalizing Human-Centered Perspectives in Explainable AI“. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. event-place: Yokohama, Japan. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3441342.

[40] Lionel Briand, Sandro Morasca und Victor R Basili. *Goal-driven definition of product metrics based on properties*. 1995.

[41] Barbara Kitchenham. „Procedures for performing systematic reviews“. In: *Keele, UK, Keele University* 33.2004 (2004), S. 1–26. DOI: 10.1016/j.infsof.2008.09.009.

[42] Rainara Maia Carvalho et al. „Quality characteristics and measures for human–computer interaction evaluation in ubiquitous systems“. In: *Software Quality Journal* 25.3 (2017), S. 743–795. DOI: 0.1145/1864349.1864395.

[43] Nava Tintarev und Judith Masthoff. „Designing and evaluating explanations for recommender systems“. In: *Recommender systems handbook*. Springer, 2011, S. 479–510. DOI: 10.1007/978-0-387-85820-3.

[44] Masahiro Sato et al. „Context style explanation for recommender systems“. In: *Journal of Information Processing* 27 (2019), S. 720–729. DOI: 10.2197/ipsjjip.27.720.

[45] Malin Eiband et al. „The Impact of Placebic Explanations on Trust in Intelligent Systems“. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA '19. event-place: Glasgow, Scotland Uk. New York, NY, USA: Association for Computing Machinery, 2019, S. 1–6. ISBN: 978-1-4503-5971-9. DOI: 10.1145/3290607.3312787.

[46] Chun-Hua Tsai und Peter Brusilovsky. „Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance“. In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. event-place: Larnaca, Cyprus. New York, NY, USA: Association for Computing Machinery, 2019, S. 22–30. ISBN: 978-1-4503-6021-0. DOI: 10.1145/3320435.3320465.

[47] Diana C. Hernandez-Bocanegra, Tim Donkers und Jürgen Ziegler. „Effects of Argumentative Explanation Types on the Perception of Review-Based Recommendations“. In: *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '20 Adjunct. event-place: Genoa, Italy. New York, NY, USA: Association for Computing Machinery, 2020, S. 219–225. ISBN: 978-1-4503-7950-2. DOI: 10.1145/3386392.3399302.

[48] James Schaffer et al. „I Can Do Better than Your AI: Expertise and Explanations“. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. IUI '19. event-place: Marina del Ray, California. New York, NY, USA: Association for Computing Machinery, 2019, S. 240–251. ISBN: 978-1-4503-6272-6. DOI: 10.1145/3301275.3302308.

[49] Katharina Weitz et al. „Do You Trust Me?: Increasing User-Trust by Integrating Virtual Agents in Explainable AI Interaction Design“. In: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. IVA '19. event-place: Paris, France. New York, NY, USA: Association for Computing Machinery, 2019, S. 7–9. ISBN: 978-1-4503-6672-4. DOI: 10.1145/3308532.3329441.

[50] Takaaki Yamada et al. „Evaluating Explanation Function in Railway Crew Rescheduling System by Think-Aloud Test“. In: *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. Kumamoto, Japan: IEEE, 2016, S. 991–994. ISBN: 978-1-4673-8985-3. DOI: 10.1109/IIAI-AAI.2016.93. (Besucht am 24. 05. 2021).

[51] Masahiro Sato, Shin Kawai und Hajime Nobuhara. „Action-Triggering Recommenders: Uplift Optimization and Persuasive Explanation“. In: *2019 International Conference on Data Mining Workshops (ICDMW)*. ISSN: 2375-9259. 2019, S. 1060–1069. DOI: 10.1109/ICDMW.2019.00155.

[52] Jacob Haspiel et al. „Explanations and Expectations: Trust Building in Automated Vehicles“. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. event-place: Chicago, IL, USA. New York, NY, USA: Association for Computing Machinery, 2018, S. 119–120. ISBN: 978-1-4503-5615-2. DOI: 10.1145/3173386.3177057.

[53] Mark Zolotas und Yiannis Demiris. „Towards Explainable Shared Control using Augmented Reality“. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. ISSN: 2153-0866. 2019, S. 3020–3026. DOI: 10.1109/IROS40897.2019.8968117.

[54] Maria Riveiro und Serge Thill. „“That’s (not) the output I expected!” On the role of end user expectations in creating explanations of AI systems“. In: *Artificial Intelligence* 298 (2021), S. 103507. ISSN: 0004-3702. DOI: 10.1016/j.artint.2021.103507.

[55] Kyle Martin et al. „Evaluating Explainability Methods Intended for Multiple Stakeholders“. In: *KI - Künstliche Intelligenz* (2021). ISSN: 1610-1987. DOI: 10.1007/s13218-020-00702-6. (Besucht am 24. 05. 2021).

[56] Chun-Hua Tsai und Peter Brusilovsky. „The effects of controllability and explainability in a social recommender system“. In: *User Modeling and User-Adapted Interaction* (2020). ISSN: 1573-1391. DOI: 10.1007/s11257-020-09281-5. (Besucht am 24. 05. 2021).

[57] Mark A. Neerincx et al. „Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance“. In: *Engineering Psychology and Cognitive Ergonomics*. Hrsg. von Don Harris. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, S. 204–214. ISBN: 978-3-319-91122-9. DOI: 10.1007/978-3-319-91122-9_18.

[58] Tim Schrills und Thomas Franke. „Color for Characters - Effects of Visual Explanations of AI on Trust and Observability“. In: *Artificial Intelligence in HCI*. Hrsg. von Helmut Degen und Lauren Reinerman-Jones. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 121–135. ISBN: 978-3-030-50334-5. DOI: 10.1007/978-3-030-50334-5_8.

[59] Ning Wang et al. „Is It My Looks? Or Something I Said? The Impact of Explanations, Embodiment, and Expectations on Trust and Performance in Human-Robot Teams“. In: *Persuasive Technology*. Hrsg. von Jaap Ham et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, S. 56–69. ISBN: 978-3-319-78978-1. DOI: 10.1007/978-3-319-78978-1_5.

[60] Lixiao Zhu und Thomas Williams. „Effects of Proactive Explanations by Robots on Human-Robot Trust“. In: *Social Robotics*. Hrsg. von Alan R. Wagner et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 85–95. ISBN: 978-3-030-62056-1. DOI: 10.1007/978-3-030-62056-1_8.

[61] Jeamin Koo et al. „Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance“. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9.4 (2015), S. 269–275. ISSN: 1955-2513, 1955-2505. DOI: 10.1007/s12008-014-0227-2. (Besucht am 26. 05. 2021).

[62] Jeamin Koo et al. „Understanding driver responses to voice alerts of autonomous car operations“. In: *International Journal of Vehicle Design* 70.4 (2016), S. 377. ISSN: 0143-3369, 1741-5314. DOI: 10.1504/IJVD.2016.076740. (Besucht am 26. 05. 2021).

[63] Hao-Fei Cheng et al. „Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders“. In: *Proceedings of the 2019 chi conference on human factors in computing systems*. 2019, S. 1–12. DOI: 10.1145/3290605.3300789.

[64] Kacper Sokol und Peter Flach. „One Explanation Does Not Fit All“. In: *KI - Künstliche Intelligenz* 34.2 (2020), S. 235–250. ISSN: 1610-1987. DOI: 10.1007/s13218-020-00637-y. (Besucht am 24. 05. 2021).

[65] Kyle Martin et al. „Developing a Catalogue of Explainability Methods to Support Expert and Non-expert Users“. In: *Artificial Intelligence XXXVI*. Hrsg. von Max Bramer und Miltos Petridis. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, S. 309–324. ISBN: 978-3-030-34885-4. DOI: 10.1007/978-3-030-34885-4_24.

[66] Upol Ehsan und Mark O. Riedl. „Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach“. In: *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*. Hrsg. von Constantine Stephanidis et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 449–466. ISBN: 978-3-030-60117-1. DOI: 10.1007/978-3-030-60117-1_33.

[67] Henrik Mucha et al. „Interfaces for Explanations in Human-AI Interaction: Proposing a Design Evaluation Approach“. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. event-place: Yokohama, Japan. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3451759.

[68] Amal Abdulrahman et al. „Belief-Based Agent Explanations to Encourage Behaviour Change“. In: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. IVA '19. event-place: Paris, France. New York, NY, USA: Association for Computing Machinery, 2019, S. 176–178. ISBN: 978-1-4503-6672-4. DOI: 10.1145/3308532.3329444.

[69] Jasper van der Waa et al. „Evaluating XAI: A comparison of rule-based and example-based explanations“. In: *Artificial Intelligence* 291 (2021), S. 103404. ISSN: 0004-3702. DOI: 10.1016/j.artint.2020.103404.

[70] Gesa Wiegand et al. „"I'd like an Explanation for That!"Exploring Reactions to Unexpected Autonomous Driving“. In: *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. Oldenburg Germany: ACM, 2020, S. 1–11. ISBN: 978-1-4503-7516-0. DOI: 10.1145/3379503.3403554. (Besucht am 24. 05. 2021).

[71] Sonja Stange und Stefan Kopp. „Effects of Referring to Robot vs. User Needs in Self-Explanations of Undesirable Robot Behavior“. In: *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '21 Companion. event-place: Boulder, CO, USA. New York, NY, USA: Association for Computing Machinery, 2021, S. 271–275. ISBN: 978-1-4503-8290-8. DOI: 10.1145/3434074.3447174.

[72] Frank Kaptein et al. „Personalised self-explanation by robots: The role of goals versus beliefs in robot-action explanation for children and adults“. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. ISSN: 1944-9437. 2017, S. 676–682. DOI: 10.1109/ROMAN.2017.8172376.

[73] Na Du et al. „Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload“. In: *Transportation research part C: emerging technologies* 104 (2019), S. 428–442. DOI: 10.1016/j.trc.2019.05.025.

[74] Jörg Cassens und Rebekah Wegener. „Ambient Explanations: Ambient Intelligence and Explainable AI“. In: *Ambient Intelligence*. Hrsg. von Ioannis Chatzigiannakis, Boris De Ruyter und Irene Mavrommati. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, S. 370–376. ISBN: 978-3-030-34255-5. DOI: 10.1007/978-3-030-34255-5_30.

[75] Douglas Cirqueira et al. „Scenario-Based Requirements Elicitation for User-Centric Explainable AI“. In: *Machine Learning and Knowledge Extraction*. Hrsg. von Andreas Holzinger et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 321–341. ISBN: 978-3-030-57321-8. DOI: 10.1007/978-3-030-57321-8_18.

[76] Khaled Rjoob et al. „Towards Explainable Artificial Intelligence and Explanation User Interfaces to Open the 'Black Box' of Automated ECG Interpretation“. In: *Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications*. Hrsg. von Thoralf Reis et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, S. 96–108. ISBN: 978-3-030-68007-7. DOI: 10.1007/978-3-030-68007-7_6.

[77] Shruthi Chari et al. „Explanation Ontology: A Model of Explanations for User-Centered AI“. In: *The Semantic Web – ISWC 2020*. Hrsg. von Jeff Z. Pan et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, S. 228–243. ISBN: 978-3-030-62466-8. DOI: 10.1007/978-3-030-62466-8_15.

[78] Mireia Ribera und Agata Lapedriza. „Can we do better explanations? A proposal of user-centered explainable AI.“ In: *IUI Workshops*. 2019.

[79] David Gunning und David Aha. „DARPA's explainable artificial intelligence (XAI) program“. In: *AI Magazine* 40.2 (2019), S. 44–58. DOI: 10.1609/aimag.v40i2.2850.

[80] Brian Y Lim und Anind K Dey. „Assessing demand for intelligibility in context-aware applications“. In: *Proceedings of the 11th international conference on Ubiquitous computing*. 2009, S. 195–204. DOI: 10.1145/1620545.1620576.

[81] Nava Tintarev und Judith Masthoff. „A survey of explanations in recommender systems“. In: *2007 IEEE 23rd international conference on data engineering workshop*. IEEE. 2007, S. 801–810. DOI: 10.1109/ICDEW.2007.4401070.

[82] Sule Anjomshoae et al. „Explainable agents and robots: Results from a systematic literature review“. In: *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*. International Foundation for Autonomous Agents und Multiagent Systems. 2019, S. 1078–1088. ISBN: 978-1-4503-6309-9.

[83] Luciana Salgado, Roberto Pereira und Isabela Gasparini. „Cultural Issues in HCI: Challenges and Opportunities“. In: *Human-Computer Interaction: Design and Evaluation*. Hrsg. von Masaaki Kurosu. Cham: Springer International Publishing, 2015, S. 60–70. ISBN: 978-3-319-20901-2. DOI: 10.1007/978-3-319-20901-2_6.

[84] Robert R Hoffman et al. *Metrics for Explainable AI: Challenges and Prospects*. 2018.

[85] High-Level Expert Group on Artificial Intelligence. *Policy and investment recommendations for trustworthy AI. A subtitle (optional)*. The European Commission, 2019.

[86] Jakob Nielsen. *Usability heuristics for user interface design*. 2010. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ (besucht am 24. 09. 2021).

[87] Reginald G Golledge et al. *Wayfinding behavior: Cognitive mapping and other spatial processes*. JHU press, 1999. ISBN: 080185993X.

[88] Ben Shneiderman et al. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016. DOI: 0321601483.

[89] Ranjana Rajnish, Prof Dev und Vyas Rajnish. „Writing Quality Requirements (SRS): An Approach To Manage Requirements Volatility“. In: *Indian Journal of Computer Science and Engineering* 1 (2010).

[90] Ian F Alexander und Richard Stevens. *Writing better requirements*. Pearson Education, 2002. ISBN: 9780321131638.

[91] Christophe Leys et al. „Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median“. In: *Journal of Experimental Social Psychology* 49 (2013), 764–766. DOI: 10.1016/j.jesp.2013.03.013.

[92] Olive Jean Dunn. „Multiple comparisons using rank sums“. In: *Technometrics* 6.3 (1964), S. 241–252. DOI: 10.1080/00401706.1964.10490181.

[93] G. Susanne Bahr und Richard A. Ford. „How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making“. In: *Computers in Human Behavior* 27.2 (2011). Web 2.0 in Travel and Tourism: Empowering and Changing the Role of Travelers, S. 776–783. ISSN: 0747-5632. DOI: 10.1016/j.chb.2010.10.030.

[94] Tony Gorschek et al. „A Model for Technology Transfer in Practice“. In: *IEEE Software* 23.6 (2006), S. 88–95. DOI: 10.1109/MS.2006.147.