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

# Evaluation von LLMs in der Anforderungspriorisierung ausgehend von Feedback zu Visionvideos

**Evaluation of LLMs in requirements prioritization based on feedback on vision videos**

## Bachelorarbeit

im Studiengang Informatik

von

**Yevgen Tytov**

**Prüfer: Prof. Dr. rer. nat. Kurt Schneider**
**Zweitprüfer: Dr. rer. nat. Jil Ann-Christin Klünder**
**Betreuer: M. Sc. Marc Herrmann**

**Hannover, 23.09.2024**

# Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit 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 23.09.2024

Yevgen Tytov

# Zusammenfassung

Requirements Engineering ist ein zentraler Prozess in der Softwareentwicklung, der die Grundlage für den Erfolg eines Projekts legt. Ein wichtiger Bestandteils davon ist die Priorisierung von Anforderungen, um sicherzustellen, dass die wichtigsten Funktionen und Bedürfnisse zuerst umgesetzt werden. Während in traditionellen Projekten Anforderungsingenieure häufig die Hauptverantwortung für die Priorisierung tragen, spielt in einem CrowdRE-Setting die aktive Beteiligung von Stakeholdern eine entscheidende Rolle. Für diese kann sich der Priorisierungsprozess jedoch als mühsam erweisen.

Eine Möglichkeit, die Anforderungspriorisierung für Stakeholder zu erleichtern, könnten Large Language Models wie ChatGPT sein, die in den letzten Jahren im privaten und gewerblichen Bereich an Popularität gewonnen haben. Diese Arbeit untersucht, ob ChatGPT Stakeholder bei der Priorisierung von Anforderungen in die Kano-Kategorien innerhalb eines CrowdRE-Settings unterstützen kann, den Ablauf beschleunigt und den Priorisierungsprozess gegenüber traditionellen Methoden erleichtert.

Die Ergebnisse der Studie zeigen, dass der Einsatz von ChatGPT den Prozess insgesamt verlängert hat. Auch eine leichte Tendenz zur Erleichterung des Priorisierungsprozesses konnte festgestellt werden, allerdings gab es weder in Bezug auf die Zeit noch auf die empfundene Schwierigkeit signifikante Unterschiede im Vergleich zu herkömmlichen Methoden. Dennoch äußerten sich die Stakeholder mit Zugang zu ChatGPT positiv darüber, dass der Chatbot eine unterstützende Funktion beim Priorisierungsprozess geboten hat.

# Abstract

Requirements Engineering is a central process in software development that lays the foundation for a project's success. A key aspect of this process is the prioritization of requirements, ensuring that the most important features and needs are implemented first. While requirements engineers often bear the primary responsibility for prioritization in traditional projects, the active involvement of stakeholders plays a crucial role in a CrowdRE setting. However, for stakeholders, this prioritization process can prove to be cumbersome.

One way to ease the prioritization of requirements for stakeholders could be through the use of large language models such as ChatGPT, which have gained popularity in both private and commercial sectors in recent years. This study examines whether ChatGPT can assist stakeholders in prioritizing requirements into Kano categories within a CrowdRE setting, accelerate the process, and simplify the prioritization compared to traditional methods.

The study's results show that using ChatGPT actually extended the overall process. There was a slight tendency towards simplifying the prioritization, but no significant differences were found in terms of time or perceived difficulty compared to conventional methods. Nonetheless, stakeholders with access to ChatGPT expressed positive views on the chatbot's supportive role in the prioritization process.

# Inhaltsverzeichnis

# Kapitel 1

# Einleitung

Requirements Engineering ist in Softwareprojekten ein fester Bestandteil des Gesamtprozesses und wird durch steigende Komplexität von Software immer bedeutender [7]. Eines der grundlegenden Bestandteile von Requirements Engineering ist die Zuordnung von Anforderungen nach Wichtigkeit, auch Anforderungspriorisierung genannt [2]. In traditionellen Projekten tragen Anforderungsingenieure die Hauptverantwortung für die Priorisierung der Anforderungen, um sicherzustellen, dass die wichtigsten Funktionen frühzeitig umgesetzt werden. Während solche herkömmliche Ansätze des Requirements Engineering oft auf eine begrenzte Anzahl von Stakeholdern zurückgreifen, z.B. durch Interviews und Fokusgruppen, ermöglicht CrowdRE (crowd based Requirements Engineering) eine große Gruppe potenzieller Nutzer und Mitwirkender in den Prozess einzubeziehen [14]. Die Priorisierung kann sich jedoch als aufwendig herausstellen, da Stakeholder oft unterschiedliche Ziele und Prioritäten verfolgen [23]. Stakeholder können so bei der Priorisierung von Anforderungen auf Herausforderungen stoßen, insbesondere wenn sie über weniger Erfahrung verfügen. Dies kann dazu führen, dass wichtige Funktionen nicht die notwendige Priorität erhalten, was den Gesamtprozess erschwert und verlängert. Ein mögliches Tool zur Unterstützung der Stakeholder könnten LLMs sein. Large Language Models (LLMs) wie ChatGPT sind vortrainierte Sprachmodelle, die natürlich Sprache verstehen [32]. Die Unterstützung und teilweise Automatisierung im Bereich des Requirements Engineering mithilfe von LLMs wurde bereits in mehreren Arbeiten erforscht [27, 25, 9]. Eine Form der Unterstützung würde so aussehen, dass der Prozess deutlich verkürzt wird. In dieser Bachelorarbeit wird untersucht, inwiefern die Anforderungspriorisierung durch LLMs übernommen werden kann. Dabei werden Stakeholder ihre Anforderungen, die mittels eines Visionvideos ermittelt werden, mithilfe ChatGPT priorisieren.

## 1.1 Problemstellung

Durch zunehmende Forderungen an eine schnelle Entwicklung im Bereich des Software Engineerings wird es immer wichtiger sich anzupassen und effektive Methoden für die Priorisierung zu finden [27]. Eine gut durchgeführte Anforderungspriorisierung spielt nicht nur eine zentrale Rolle für die Einhaltung von Budgets und Zeitplänen, sondern ist auch entscheidend für den erfolgreichen und pünktlichen Release von Softwareversionen [2]. Das kann sich aber als schwierig und Zeitintensiv erweisen. Dieser Prozess birgt jedoch Fehler, da neue, visionäre Projekte das Priorisieren durch mangelnde Erfahrung der Stakeholder erschweren kann.

Das führt zu Verzögerung, kann die Qualität des Produkts reduzieren, Kosten in späteren Iterationen erhöhen oder führt sogar zum scheitern des Projekts [4]. Selbst wenn Anforderungen gründlich ausgewählt und priorisiert werden, bleibt der Prozess zeitaufwendig und für Stakeholder oft als schwierig empfunden. Durch den Einsatz von unterstützenden Technologien wie Large Language Models könnte es möglich sein, den Priorisierungsprozess zu beschleunigen und für Stakeholder zugänglicher zu gestalten.

## 1.2 Lösungsansatz

In dieser Arbeit wird eine Studie durchgeführt. Ziel dieser Studie ist es, die Frage zu beantworten, ob LLMs für die Anforderungspriorisierung eingesetzt werden können. Die Stakeholder werden anfangs mittels eines Online Surveys zu Visionvideos Feedback geben. Nach Karras et al. [16] sind Visionvideos Videos, die eine Vision oder Teile davon darstellen, um ein gemeinsames Verständnis für ein zukünftiges System zwischen allen Beteiligten zu erreichen. Beteiligte sind in diesem Fall Stakeholder des zukünftigen Systems, sowie das Projektteam, das die Vision implementiert. Das Feedback wird anschließend vom Moderator der Studie, in dem Fall vom Autor dieser Arbeit, in Anforderungen strukturiert. Die Aufgabe der Studie besteht darin, die Anforderungen in die jeweiligen Kano-Kategorien zu priorisieren. Das Kano-Modell wurde von Noriaki Kano eingeführt und beschreibt fünf Kategorien, in die Anforderungen eingeordnet werden können [19]. Der nächste Schritt ist die Aufteilung der Stakeholder in zwei Gruppen. Um die Aufgaben im Priorisierungsprozess zu lösen, wird die Hälfte der Stakeholder ChatGPT nutzen, während sich die andere Hälfte nur herkömmlichen Methoden bedient. Abschließend werden die gewonnen Daten gesammelt, ausgewertet und interpretiert.

Folgende Forschungsfragen werden beantwortet:

**[RQ1]**
*Wie unterscheidet sich die Geschwindigkeit der Priorisierung von Anforderungen mit LLMs in Vergleich zu ohne?*

Wir möchten die Zeit beider Gruppen messen und vergleichen. Auch wenn die Studie nicht einer realen Umgebung entspricht, werden die Erkenntnisse der Forschungsfrage eine grobe Richtung vorgeben, ob sich der Einsatz von LLMs allein zeitlich überhaupt lohnt.

Die Zeit wird während der Studie erfasst und anschließend dokumentiert.

**[RQ2]**
*Wie unterscheidet sich die Schwierigkeit der Anforderungspriorisierung mit LLMs in Vergleich zu ohne?*

Wie bereits in der Einleitung erfasst, ist die Priorisierung für Stakeholder eine Herausforderung. Diese Forschungsfrage untersucht, ob das Nutzen von LLMs die empfundene Schwierigkeit senkt und die Stakeholder entlasten kann. Zur Schwierigkeit zählt das Verstehen und Anwenden von Priorisierung in die jeweiligen Kano-Kategorien, sowie das Finden einer Begründung (Rationale) für die Priorisierung. Die Gruppen werden hier jeweils ihre eigenen Herausforderungen zu meistern haben. während die LLM-Gruppe sich mit ChatGPT und dem richtigen Promt vertraut machen muss, müssen die anderen Stakeholder das Priorisieren in die Kano-Kategorien ohne die Hilfe von ChatGPT verstehen, diese händisch durchführen und eigenständig eine Begründung finden.

Um die notwendigen Daten zu erfassen, wird am Ende der Studie eine Umfrage durchgeführt.

**[RQ3]**
*Kann ChatGPT Stakeholder beim priorisieren von Anforderungen unterstützen?*

Mithilfe eines Fragebogen werden die Stakeholder der ChatGPT-Gruppe nach der Priorisierung aller Anforderungen angeben, ob ihnen ChatGPT beim Prozess helfen konnte. Genauer wird gefragt, inwiefern ChatGPT in folgenden Bereichen hilfreich ist: 1.: das Kano-Modell besser zu verstehen, 2.: Priorisierung von Anforderungen in die Kano-Kategorien, 3.: Begründung für die Priorisierung formulieren. Des weiteren wird untersucht, zu welchem Grad die Priorisierungen von ChatGPT nachvollziehbar sind.

## 1.3 Struktur der Arbeit

Im folgenden wird die Struktur der Arbeit zusammengefasst.

In Kapitel 2 wird auf die für diese Arbeit benötigten Grundlagen von LLMs eingegangen. Zusätzlich werden die für die Studie relevanten Themen CrowdRE und Kano-Kategorien kurz erklärt. Im Anschluss wird in Kapitel 3 auf verwandte Arbeiten eingegangen und zu dieser Arbeit abgegrenzt. In Kapitel 4 wird der Aufbau und die Durchführung der Studie erläutert. Außerdem werden wichtige Details zu den Teilnehmern und für die Studie wichtige Umfrage erwähnt. Das Kapitel 5 befasst sich mit den Ergebnissen der Studie. Die Interpretation der Ergebnisse, die Beantwortung der Forschungsfragen und den Einschränkungen dieser Arbeit widmet sich Kapitel 6. Abschließend wird in Kapitel 7 die Arbeit zusammengefasst und einen Ausblick für zukünftige Arbeiten vorgestellt.

# Kapitel 2

# Grundlagen

In diesem Kapitel werden Verfahren und Technologien erklärt, die für das Verständnis dieser Arbeit notwendig sind. Erst werden Large Language Models, insbesondere ChatGPT erläutert. Anschließend werden die Themen CrowdRE, Kano-Kategorien und das FunktionsMASTER-Modell kurz beschrieben.

## 2.1 Large Language Model

Large Language Models (LLMs) sind vortrainierte, KI-unterstützte Sprachmodelle, mit teilweise über hunderten Milliarden an Parametern, die menschliche Sprache verstehen, verarbeiten und generieren können. Dadurch können LLMs in vielen verschiedenen Anwendungsbereichen, wie die Generierung von Software Requirements Spezifikation und Code, eingesetzt werden [32, 17]. Hierbei ist die Anzahl der Parameter ein entscheidender Faktor, dass ein „Language Model“ als „Large Language Model“ kategorisiert wird, da dadurch das Leistungsniveau signifikant ansteigt und sie besondere Fähigkeiten wie kontextunabhängiges Lernen aufweisen [34]. Um die Forschungsfragen dieser Arbeit zu beantworten, wird ChatGPT genutzt. ChatGPT, entwickelt von OpenAI, ist ein Chat-basiertes Large Language Model, das besonders gut für die Verarbeitung von natürlicher Sprache geeignet ist [24]. ChatGPT bietet zum Zeitpunkt der Veröffentlichung dieser Arbeit zwei Versionen: GPT-4o und GPT-4o mini. Während GPT-4o das neuste und fortschrittlichste Modell von OpenAI ist, ist der Zugang kostenpflichtig. GPT-4o mini hingegen ist kostenlos, intelligenter als GPT-3.5-turbo und gut für kleinere Aufgaben geeignet [1]. Um jeden Stakeholder unbeschränkten Zugriff auf ChatGPT zu ermöglichen, wird für diese Arbeit GPT-4o mini verwendet.

Um mit ChatGPT zu interagieren, wird eine Nachricht in natürlicher Sprache verfasst und abgeschickt. Diese Nachricht wird *prompt* genannt und beeinflusst die Ausgabe von ChatGPT, wobei die Qualität der Ausgabe durch Prompt Engineering erhöht werden kann. Prompts sind also der

Schlüssel, um Large Language Models dazu zu bringen, nützliche und relevante Ergebnisse zu liefern [21]. In dieser Arbeit haben Stakeholder die Möglichkeit, die Qualität der Ergebnisse von ChatGPT mithilfe einer Rollenzuteilung, der Angabe des Kontextes und einer genauen Beschreibung der Aufgabe potenziell zu erhöhen. Hierbei kann eine zugewiesene Rolle den Effekt erzielen, dass ChatGPT eine Aufgabe so erledigt, wie es auch ein Mensch, also zum Beispiel ein Anforderungsingenieur, tun würde. Zu beachten ist, dass solche Vorbereitungen idealerweise der erste Prompt des Chats sind [15].

Ein beispielhafter Prompt, um die Ergebnisse für eine Priorisierung von Anforderung potenziell zu verbessern, könnte lauten: „Nimm ab jetzt die Rolle eines Anforderungsingenieurs ein. Deine Aufgabe wird sein, Anforderungen nach Kano in ein Basismerkmal, Leistungsmerkmal oder Begeisterungsmerkmal zu priorisieren und anschließend zu begründen. Die Anforderungen sind für ein Augmented Reality System eines Autos.“. Dieser Prompt gibt ChatGPT sowohl eine Rolle, ein Kontext und eine Beschreibung für die bevorstehende Aufgabe an.

Korrektur von Ergebnissen: LLMs wie ChatGPT sollen die Entscheidungsfindung für die Priorisierung von Anforderungen erleichtern und Stakeholder dabei unterstützen, sodass Anforderungsingenieure entlastet oder bei dieser Tätigkeit sogar komplett ersetzt werden können. Da es jedoch wichtig ist, Anforderungen nach der Sicht der Stakeholder zu priorisieren, können diese ChatGPT's Ergebnisse anzweifeln und nach alternativen Lösungen fragen. Für diese Arbeit wird die Möglichkeit, dass ChatGPT seine Entscheidung überdenkt, durch Prompt Engineering und der „Regenerate-Funktion“ bereitgestellt. Letzteres ist eine Funktion von ChatGPT, Nachrichten neu zu generieren.

## 2.2 CrowdRE

CrowdRE steht für Crowd-based Requirements Engineering und ist ein Oberbegriff für (halb)automatisierte Ansätze zum sammeln und Analysieren von Informationen aus einer Crowd. Mit einer Crowd sind hier Stakeholder, potenzielle Stakeholder und Menschen gemeint, die vom Produkt profitieren [14]. In dieser Arbeit werden so durch eine Crowd Anforderungen an einem Visionvideo eines Augmented-Reality-Systems eines Autos erhoben und zu einem späteren Zeitpunkt nach dem Kano-Modell priorisiert. Stakeholder können dabei potenzielle Nutzer des AR-Systems des Autos sein, oder auch Passanten, die durch das Produkt potenziell sicherer sind.

## 2.3 Modell FunktionsMASTER

Um die aus dem Visionvideo extrahierten Anforderungen einheitlich zu formulieren, wurde das FunktionsMASTER-Modell für funktionale- und nicht-funktionale Anforderungen verwendet. FunktionsMASTER ist ein schablonenbasierter Ansatz, um Aufwände beim Ermitteln, Vermitteln und Dokumentieren von unter anderem Anforderungen zu minimieren [26].

Bevor auf die genutzten Schablonen eingegangen wird, folgt hier eine kurze Erklärung von funktionalen und nicht-funktionalen Anforderungen. Eine funktionale Anforderung beschreibt, was ein Produkt oder ein System tun soll. Es charakterisiert also das Verhalten eines Systems [13]. Nicht-funktionale Anforderungen hingegen beschreiben nicht das Was, sondern das Wie, genauer gesagt die Eigenschaften, Merkmale oder Einschränkungen, die ein (Software)System ausweisen muss [20].

Die Anforderungsschablone in Abbildung 2.1 hilft, funktionale Anforderungen ohne Bedingungen zu formulieren. Hierbei werden, wie bei allen drei Schablonen, aufsteigend alle Schritte abgearbeitet und lediglich die Werte in <> ersetzt. Ein Beispiel dazu könnte lauten: „Das Augmented-Reality-System im Auto muss fähig sein, die Daten des Navigationssystems zu empfangen“. Hier wäre der Betrachtungsgegenstand das AR-System, die Wichtigkeit „muss“, die Funktionalität „empfangen“, die Art der Funktionalität „fähig sein“ und das Objekt, für das die Funktionalität gefordert wird, „Daten des Navigationssystems“.

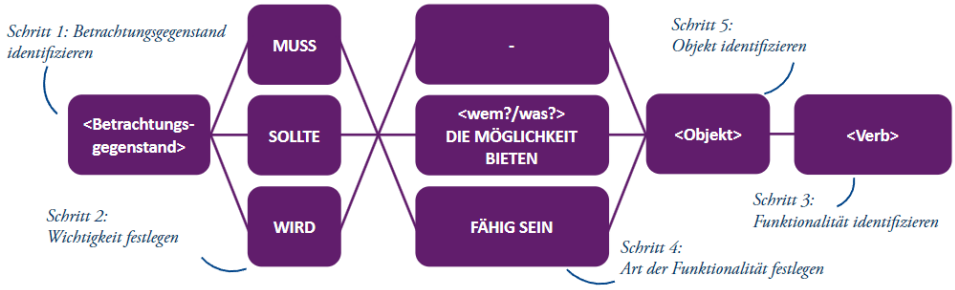



Abbildung 2.1: Funktional ohne Bedingung [26]

Die Anforderungsschablone in Abbildung 2.2 für funktionale Anforderungen mit Bedingen wird ähnlich wie die in Abbildung 2.1 verwendet. Einzig die Bedingung, die neu als Schritt sechs hinzugefügt wurde und der verschobene Betrachtungsgegenstand sind anders, wodurch die etwas anders strukturierte Anforderung jetzt mit einer Bedingung formuliert werden kann. Ein Beispiel könnte lauten: „Sobald sich das Auto innerhalb von 300 Metern eines interessanten Orts befindet, sollte das Augmented-Reality-System des Autos fähig sein, auf der Windschutzscheibe interessante Orte zu markieren“.

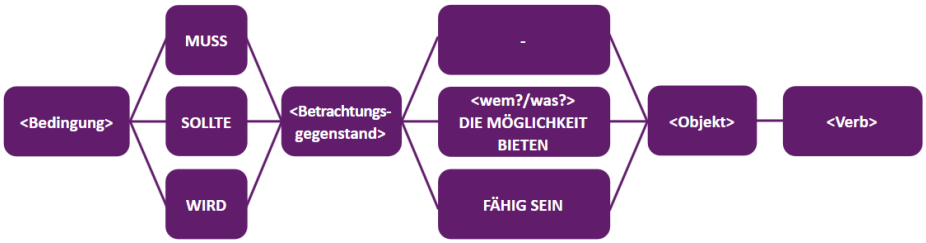



Abbildung 2.2: Funktional mit Bedingung [26]

Für nicht-funktionale Anforderungen wird in dieser Arbeit die letzte Schablone in Abbildung 2.3 verwendet. Neu sind die Knoten „Eigenschaft“, die den Betrachtungsgegenstand charakterisiert, der optionale „Vergleichsausdruck“ und der „Wert“ der Eigenschaft. Ein Beispiel für eine nicht-funktionale Eigenschaft wäre: „Das Augmented-Reality-System des Autos muss eine Reaktionszeit von 10ms garantieren“, wobei hier die Eigenschaft die Reaktionszeit darstellt und der Wert 10ms ist.

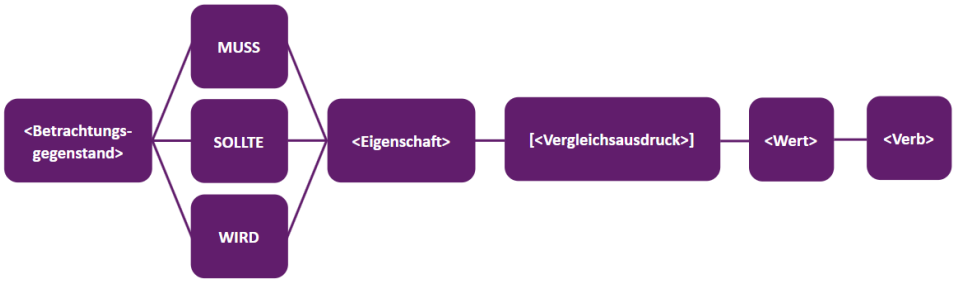



Abbildung 2.3: Nichtfunktional [26]

Aufgrund des Studiendesign, dass Stakeholder bestehende Anforderungen ändern, löschen und neue Anforderungen hinzufügen konnten, sind nur die Basisanforderungen, die vom Autor dieser Arbeit erstellt worden sind, mit dem Modell verfasst worden. Das stellt hier jedoch kein Problem dar, weil, wie bereits erwähnt, das Modell lediglich der Vereinheitlichung der Umfrage diente und bei der Priorisierung später, aufgrund der variablen Anforderungsanzahl, kein uniformes Formular gab.

## 2.4 Kano-Modell

Das im Jahre 1984 von Noriaki Kano erstellte Kano-Modell stellt eine Methodik dar, um Anforderungen von Kunden nach Wichtigkeit zu beschreiben [19]. Spätere Studien konnten basierend darauf zeigen, dass

das Kano-Modell deshalb auch zum Priorisieren von Kundenanforderungen geeignet ist. Das Modell basiert auf der two-factor-Theorie, die besagt, dass ein Faktor, der Zufriedenheit schafft, bei Abwesenheit nicht zwingend zur Unzufriedenheit führt [6]. Nach Kano kann die Kundenzufriedenheit mit einer nichtlinearen Funktion der Produktfunktionalität beschrieben werden, wobei sich diese aus drei Hauptkomponenten zusammensetzt: „muss“, „soll“ und „kann“. Das Muss-Merkmal (oder Basismerkmal) stellt Anforderungen dar, die Kunden als selbstverständlich ansehen. Kann ein Produkt dieses Merkmal vorweisen, entsteht keine Zufriedenheit. Fehlt dieses Merkmal jedoch, entsteht durch die Selbstverständlichkeit Unzufriedenheit. Diese Kategorie von Anforderungen repräsentieren ein Mindestniveau des Produkts oder der Dienstleistung [19]. Das Soll-Merkmal (auch Leistungsmerkmal) sind eindimensionale Anforderungen, die bei Vorhandensein Zufriedenheit schaffen, bei Abwesenheit jedoch Unzufriedenheit verursachen. Schlussendlich repräsentiert das Kann-Merkmal (Begeisterungsmerkmal) „attraktive“ Anforderungen, die Kunden weder erwarten noch explizit äußern. In diesem Fall schafft Vorhandensein Zufriedenheit, die Abwesenheit dieser Merkmale jedoch keine Unzufriedenheit [18]. Zur Vollständigkeit sei hier erwähnt, dass das Kano-Modell in vielen Publikationen mit fünf statt drei Merkmalen behandelt wird, was jedoch für diese Arbeit nicht relevant ist, da nur die ersten drei hier erwähnten Merkmale beachtet werden.

Um die Anforderungen in Kategorien einzuteilen, wird standardgemäß ein Fragebogen verwendet. In diesem werden für jede Anforderung zwei verschiedene Fragen gestellt: eine funktionale und eine dysfunktionale. Die funktionale Frage soll die Zufriedenheit des Kunden bewerten, wenn die Anforderung erfüllt ist. Die dysfunktionale Frage bewertet hingegen die Zufriedenheit bei Nichterfüllung der Anforderung [19]. Die beiden Fragen können dann jeweils, wie in Abbildung 2.4 zu sehen, auf fünf verschiedene Weise beantwortet werden. Der ausgefüllten Fragebogen wird schließlich so ausgewertet, dass beide Antworten jeder Anforderung betrachtet wird und so ein Merkmal abgeleitet werden kann. Eine Anforderung, die Beispielsweise funktional ein „It must be that way“ und dysfunktional ein „I dislike it that way“ erhalten hat, wird als ein Basismerkmal klassifiziert. Der Kunde ist in diesem Beispiel der Meinung, dass die Anforderung erfüllt sein muss (funktional) und eine Abwesenheit der Anforderung stören würde (dysfunktional).

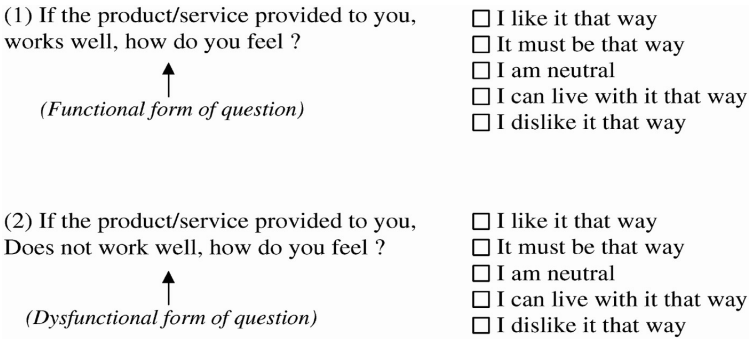



Abbildung 2.4: Kano evaluation table [28]

Um das Vorgehen zu vereinfachen, enthält der Fragebogen zur Priorisierung keine Funktionalen und dysfunktionalen Fragen, sondern nur die Kano-Merkmale. Stakeholder werden über die jeweiligen Kano-Merkmale informiert und Beispiele gezeigt, wobei Verständnisfragen beantwortet werden. Anschließend sollen beide Gruppen (ohne und mit ChatGPT) die Anforderungen in die jeweilige Kategorie direkt einordnen und die Entscheidung begründen.

# Kapitel 3

# Verwandte Arbeiten

Dieser Abschnitt stellt verwandte Forschungsarbeiten vor, die sich, im Kontext von Requirements Engineering, mit Anforderungspriorisierung, Visionvideos und (KI-unterstützte) Automatisierung von Anforderungen befasst haben. Anschließend wird diese Arbeit zu den gezeigten verwandten Arbeiten abgegrenzt.

## 3.1 Visionvideos

In der Bachelorarbeit von Dryaev [9] wurde untersucht, ob LLMs bei der Erstellung und Verbesserung von Anforderungen aus Visionvideos unterstützen können. Hierbei schreiben Stakeholder Kommentare zu einer Spezifikation, woraufhin auf Basis der Kommentare die Spezifikation verbessert werden soll. Mit Verbesserung ist gemeint, dass ein Anforderungsingenieur das Ergebnis von ChatGPT regenerieren kann, damit die Spezifikation nach dem Kommentar des Stakeholders dementsprechend angepasst wird. Das Ergebnis der Studie, die mit Stakeholdern durchgeführt wurde, zeigte, dass ChatGPT auf einem Niveau mit den Manuellen sind. Es wurde beobachtet, dass beim Nutzen von ChatGPT sowohl die Effizienz steigt, als auch die Anzahl von Syntax- und Formulierungsfehlern. ChatGPT kann mit manuellen Verbesserungen von Anforderungen mithalten, ein Anforderungsingenieur kann es aber nicht ersetzen, da diese die Ergebnisse überprüfen müssen

In der Bachelorarbeit von Yang [31] wurde der Zusammenhang von Emotionen von Stakeholder und der Priorisierung von Qualitätsanforderungen untersucht. In einer Nutzerstudie schauten sich Teilnehmende Visionvideos an und teilten ihre Emotionen bezüglich der gezeigten Qualitätsanforderungen mit. Anschließend wurde mittels eines Interviews die Priorisierung der Anforderungen vorgenommen. Mittels einer Korrelationsanalyse konnte festgestellt werden, dass bei von den Stakeholdern hoch priorisierten Anforderungen die verbundenen Emotionen stärker waren als bei den niedrig priorisierten Anforderungen.

Piam [22] hat in seiner Masterarbeit untersucht, ob es mithilfe von der Sprache Gherkin Stakeholdern leichter fällt, Anforderungen aus Vision Videos zu erheben. Gherkin ist eine natürliche Sprache, um Anforderungen anwendungsnah zu formulieren. Dazu wurden die Visionvideos mithilfe von Gherkin textuell beschrieben und anschließend Anforderungen abgeleitet. Durch eine Studie konnte festgestellt werden, dass die Stakeholder, die Gherkin nutzten, mehr Anforderungen formulieren konnten als ohne. Die Schlussfolgerung zu diesem Ergebnis ist, dass Anforderungen mithilfe von Gherkin für Stakeholder leichter zu erkennen sind.

## 3.2 LLMs im Requirements Engineering

Im Paper von Fantechi et al. [12] wurde untersucht, wie zuverlässig ChatGPT Inkonsistenz in Anforderungen finden kann. Im Experiment wurden bestehende Anforderungsdokumente verwendet und widersprüchliche Anforderungen hinzugefügt. Die Ergebnisse von ChatGPT wurde mit denen von Experten verglichen, die ebenfalls Beispiel-Anforderungen auf Inkonsistenz prüften. Das Experiment zeigte, dass, obwohl kostenintensiver, manuelle Überprüfung bessere Ergebnisse als ChatGPT liefert.

Zhang et al. [33] haben das tool PersonaGen entwickelt, das mithilfe von GPT-4 und knowledge graphs Personas erstellt und Anforderungsanalyse in agiler Softwareentwicklung erleichtern soll. PeronaGen bekommt Nutzerfeedback als input und generiert daraufhin eine Persona, die helfen soll, die Nutzergruppe des Produkts besser zu Verstehen. Durch eine Studie konnte gezeigt werden, dass mithilfe des tools funktionale Anforderungen leichter adressiert werden konnten, während nicht-funktionelle Anforderungen eine Herausforderung darstellten.

In der Arbeit von Chen et al. [8] wird untersucht, wie GPT-4 beim modellbasierten Anforderungsmanagement eine Hilfe sein kann. Der Fokus liegt hier bei zielorientierten Modellen. Beim modellbasierten Anforderungsmanagement werden Anforderungen nicht in natürlicher Sprache verfasst, sondern als grafisches Modell dargestellt. Dafür wird hier die Goal-orientend Requirement Language (GRL) verwendet. Durch eine Fallstudie konnte gezeigt werden, dass viele von GPT-4 generierten Ergebnisse generisch oder sogar falsch waren. Es wurde außerdem beobachtet, dass mehrere Durchläufe bessere Ergebnisse erzielten als ein einzelner Durchlauf. Die Autoren sehen trotzdem einen Wert in LLMs, unter anderem für Stakeholder mit wenig Expertise in diesem Umfeld, da GPT-4 nützliche Ideen aufmerksam machen kann, die für Stakeholder außerhalb des Fachgebiets nicht offensichtlich sind.

Im Paper von Ruan et al. [25] wird ein automatisiertes Framework vorgestellt, das mithilfe von ChatGPT Anforderungsmodelle generiert. Hierbei werden Informationen aus Anforderungstexten extrahiert und anhand vordefinierter Regeln wieder zusammengesetzt. Die von ChatGPT konstruierten

Anforderungsmodelle wurden mit manuell erstellten Modellen verglichen. Das Ergebnis war, dass generierte Modelle zwar schnell und leicht zu erstellen sind, aber ein Experte, der diese Modelle prüft und vervollständigt, immer noch unumgänglich ist.

In der Arbeit von Xie et al. [30] wird mittels einer empirischen Studie das Potenzial von LLMs bei der Generierung von Software Spezifikationen untersucht. Dafür wurden 15 LLMs getestet und mit herkömmlichen Techniken verglichen. Die Ergebnisse der Studie zeigten, dass LLMs mittels Few-Shot learning und dem richtigen prompt traditionelle Methoden übertreffen können. Dennoch leiden die Resultate von LLMs bei ineffektiven Prompts und fehlendem Domänenwissen.

## 3.3 Anforderungspriorisierung

Sami et al. [27] stellen in ihrem Paper ein webbasiertes tool vor, das LLMs und prompt engineering einsetzt, um Anforderungspriorisierung zu automatisieren und dabei verschiedene Priorisierungstechniken anwendet. Das tool funktioniert so, dass ein Nutzer seine Anforderung in Klartext mit der gewünschten Anzahl an user stories eingibt, eine Art der Priorisierung wählt (z.B. MoSCoW) und die LLM anschließend die Priorisierung vornimmt und ausgibt.

In der Arbeit von Duan et al. [10] wird ein Ansatz zur Automatisierung von einem Großteil des Anforderungspriorisierungsprozesses beschrieben, um Stakeholder bei mühsamen Aufgaben zu entlasten. Die Methode nutzt data-mining und Machine Learning, um Anforderungen nach Stakeholderinteressen, Geschäftszielen und Anliegen wie Sicherheit und Leistung zu priorisieren. Die Schlussfolgerung nach zwei Fallstudien war, dass die Automatisierung hilfreich ist. Das volle Potenzial entfaltet sich jedoch eher bei großen Projekten mit tausenden von unstrukturierten Anforderungen.

Binder et al. [6] untersuchten in ihrer Arbeit einen Ansatz, um App-Bewertungen nach den Faktoren des Kano-Modells zu klassifizieren. Zu den Faktoren gehören 'basic needs', 'performance factors', 'delighters' und 'irrelevent'. Für das Training der classifiere, welche die Automatisierung übernehmen sollen, wurden zwei Datasets mit tausenden von App-reviews verwendet. In dieser Arbeit wurden vier classifiere getestet, die jeweils ein anderes Language Model nutzten. Dazu gehörten BERT, RoBERTa, RemBERT und ALBERT. Das Klassifizieren funktioniert so, dass bei den Bewertungen nach Schlüsselwörtern gesucht wird, um diese dann in die vorgesehene Kano-Faktoren einzuordnen. Damit der classifiere den richtigen Kano-Faktor findet, wurde die Beschreibung der vier Faktoren im vorhinein manuell getätigt. Die Ergebnisse zeigten, dass eine Automatisierung, vor allem mit BERT, viel Potenzial beinhaltet. Im Vergleich zur manuellen Klassifizierung mit dem Kano-Modell ist der automatische Ansatz kosten-

günstiger und scaled besser mit großen Datenmengen an Feedback. Die Autoren denken, dass das Requirements Engineering dadurch unterstützt werden kann.

## 3.4 Abgrenzung

In dieser Arbeit wird Anforderungspriorisierung mithilfe von ChatGPT untersucht und mit traditioneller Priorisierung verglichen. Dafür werden Stakeholder zu einem Visionvideo Feedback geben, welches die Grundlage für die zu priorisierenden Anforderungen darstellt. In einer darauffolgenden Studie werden dieselben Stakeholder ihre eigenen Anforderungen in Kano-Kategorien priorisieren, wobei eine Gruppe ChatGPT nutzt, während der anderen Gruppe nur handelsübliche Mittel zur Verfügung stehen. Zwar existieren bereits ähnliche Arbeiten zu dem Thema, jedoch meines Wissens nach keine, die mit Visionvideos Anforderungen erheben und eine Studie im CrowdRE-setting durchführen, um diese Anforderungen schließlich von Stakeholdern in Kano-Kategorien zu sortieren.

# Kapitel 4

# Studiendesign

## 4.1 Goal Question Metric

Um Ziele der Studie festzulegen und die Forschungsfragen mit den richtigen Metriken zu beantworten, wird das Goal Question Metric (GQM) Verfahren nach Basili et al. [5] verwendet.

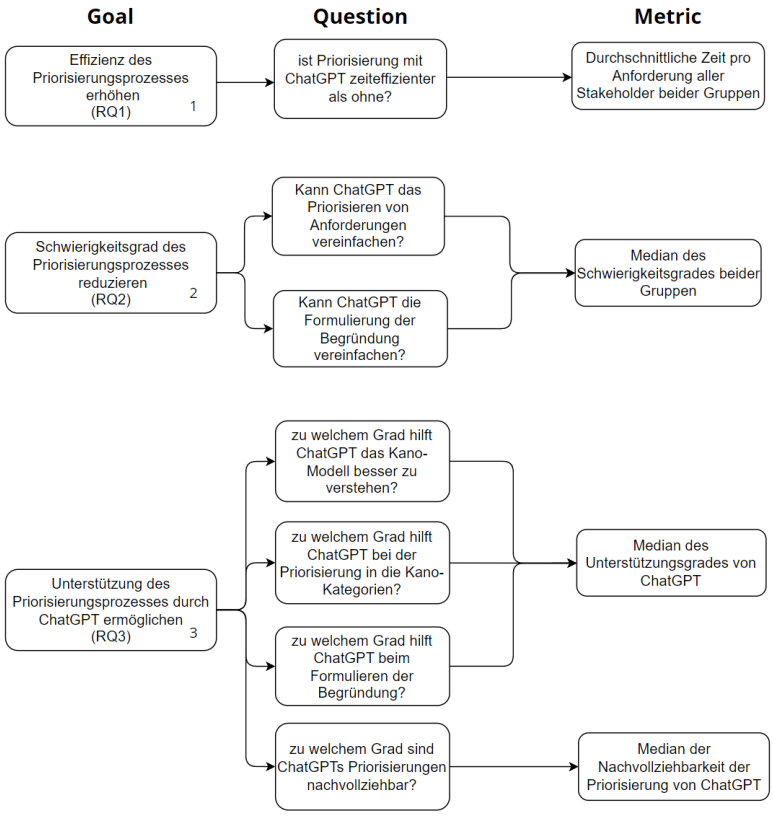



Abbildung 4.1: Goal Question Metric

Die drei Ziele der GQM werden hier mit dem Template nach Wohlin et al.[29] genauer verdeutlicht.

> **Ziel 1**
> Wir *analysieren* die Priorisierung von Anforderungen in die Kano-Kategorien mit und ohne ChatGPT,
> *zum Zweck* der Erkenntnisgewinnung, ob die Priorisierung mithilfe von ChatGPT schneller ist,
> *in Bezug auf* Effizienz,
> *vom Blickwinkel* eines Stakeholders,
> *im Kontext* eines Priorisierungsexperiments.

> **Ziel 2**
> Wir *analysieren* die Priorisierung von Anforderungen in die Kano-Kategorien mit und ohne ChatGPT,
> *zum Zweck* der Erkenntnisgewinnung, ob die Priorisierung durch ChatGPT erleichtert werden kann,
> *in Bezug auf* die empfundene Schwierigkeit,
> *vom Blickwinkel* eines Stakeholders,
> *im Kontext* eines Priorisierungsexperiments.

> **Ziel 3**
> Wir *analysieren* die Priorisierung von Anforderungen in die Kano-Kategorien mit ChatGPT,
> *zum Zweck* der Erkenntnisgewinnung, ob ChatGPT beim Priorisierungsprozess unterstützen kann,
> *in Bezug auf* den Unterstützungsgrad,
> *vom Blickwinkel* eines Stakeholders,
> *im Kontext* eines Priorisierungsexperiments.

Für die Ziele 1 und 2 werden beide Gruppen betrachtet, das Ziel 3 beschäftigt sich jedoch ausschließlich mit der ChatGPT-Gruppe. Ebenfalls werden die Fragen von Ziel 2 und 3 mithilfe einer 5-stufigen Likert-Skala beantwortet, während für die Frage von Ziel 1 die Zeit gemessen wird. Die Ankreuzmöglichkeiten der Likert-Skalen sind identisch und lauten: *Voll und ganz*, *Ein wenig*, *neutral*, *kaum*, *Gar nicht*.

## 4.2 Hypothese

In dieser Sektion werden die Hypothesen dieser Arbeit aufgestellt, wobei sich H1 und H2 jeweils auf Frage 1 und 2 der GQM beziehen.

> **Hypothese *H*1**
> Es gibt einen signifikanten Unterschied zwischen der für die Priorisierung durchschnittlich benötigten Zeit pro Anforderung

beider Gruppen.
**Nullhypothese $NH1$**
Es gibt keinen signifikanten Unterschied zwischen der für die Priorisierung durchschnittlich benötigten Zeit pro Anforderung beider Gruppen.

**Hypothese $H2_1$**
Es gibt einen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Priorisieren von Anforderungen.
**Nullhypothese $NH2_1$**
Es gibt keinen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Priorisieren von Anforderungen.

**Hypothese $H2_2$**
Es gibt einen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Formulieren einer Begründung für die Priorisierung.
**Nullhypothese $NH2_1$**
Es gibt keinen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Formulieren einer Begründung für die Priorisierung.

**Hypothese $H_{Kano}$**
Es gibt einen signifikanten Unterschied zwischen dem Verständnis beider Gruppen bezüglich des Kano-Modells.
**Nullhypothese $NH_{Kano}$**
Es gibt keinen signifikanten Unterschied zwischen dem Verständnis beider Gruppen bezüglich des Kano-Modells.

## 4.3 Variablen

Für die Studie werden folgendermaßen die abhängigen und unabhängigen Variablen definiert:
Die unabhängige Variable ist die Nutzung von *ChatGPT*, weil diese bewusst manipuliert wird, sodass die Hälfte der TeilnehmerInnen ChatGPT nutzen dürfen und die andere Hälfte nicht.
Die abhängigen Variablen verändert sich abhängig von der Wahl der unabhängigen Variable. Demnach ist die benötigte *Zeit* und die *empfundene*

*Schwierigkeit* abhängig von der Entscheidung, ob ChatGPT verwendet werden darf.

## 4.4 Teilnehmer

Die Auswahl der Probanden wurde mittels Convenience Sampling durchgeführt, wobei lediglich Minderjährige von der Teilnahme ausgeschlossen waren. Bei dieser nicht-probabilistischen Methode werden Teilnehmer ausgewählt, die aufgrund ihrer geografische Nähe oder leichten Erreichbarkeit verfügbar sind [11]. Es wurden verschiedene Studierende aus mehreren Fachbereichen der Leibniz Universität Hannover, die zum Zeitpunkt eine Abschlussarbeit an diesem Fachgebiet schreiben, per Email angeschrieben. Durch das zusätzliche Anwerben von Familie, Freunde und Bekannte konnte die Stichprobe an Stakeholdern erweitert werden, da sich hier weniger Studierende befanden und dafür mehr Arbeitnehmer.

## 4.5 Durchführung/Ablauf

Der in Abbildung 4.2 gezeigte Ablauf der Studie wurde vollständig online durchführt und bestand aus drei Schritten: die Vorbereitung, eine Umfrage und eine Priorisierung von Anforderungen mit einem kurzen Fragenteil. Ziel der Studie war es, die Unterschiede von Priorisierung mit und ohne ChatGPT zu erfassen, wobei die Umfrage lediglich zur Erfassung von demografischen Daten und dem Erstellen von Anforderungen mithilfe eines Visionvideos diente.

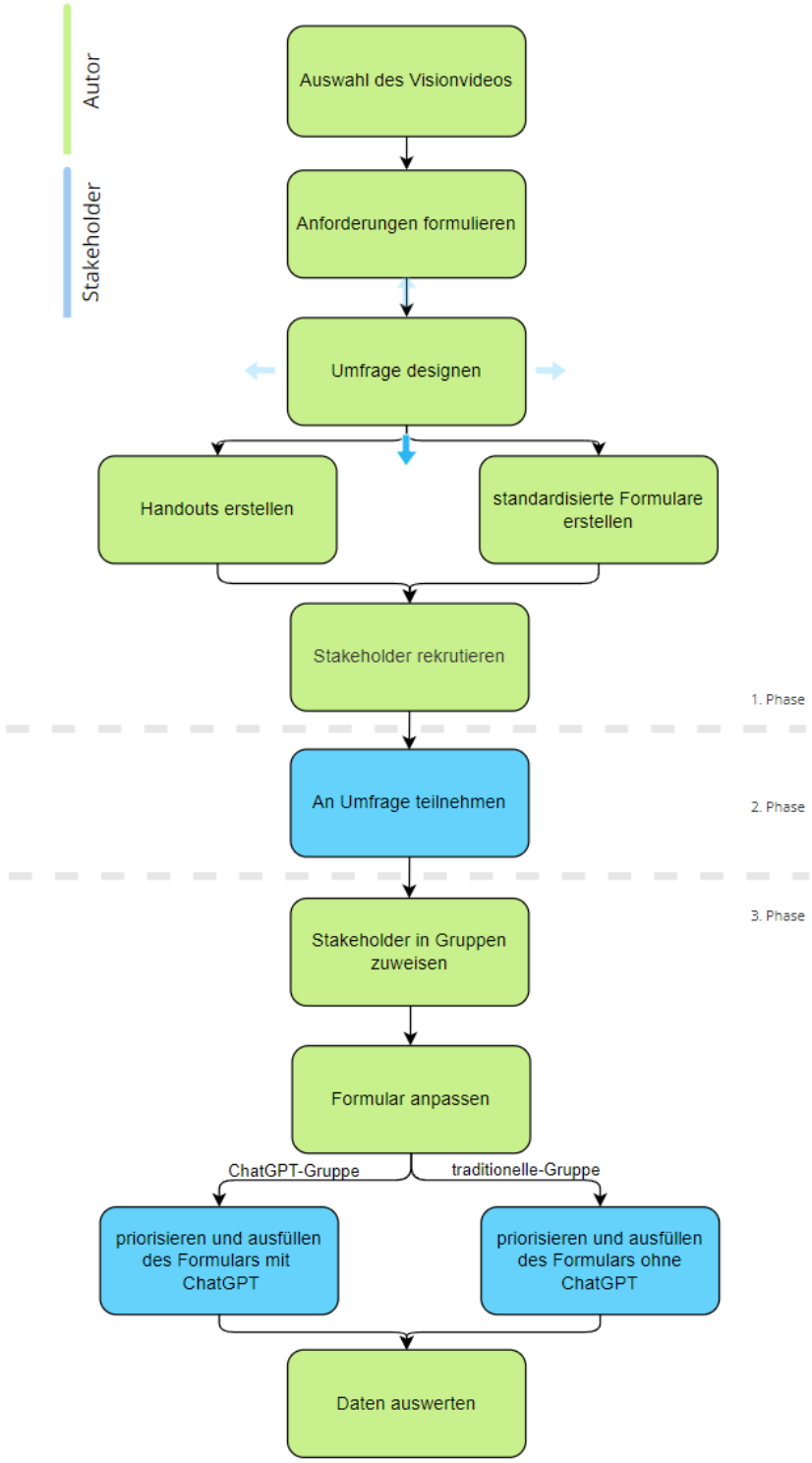



Abbildung 4.2: Verlauf der Studie

### 4.5.1 Vorbereitung

Vor Beginn der Studie wurde aus einer Menge von Visionvideos eins zur Erfassung von Anforderungen ausgewählt. Dabei wurde darauf geachtet, dass die zukünftigen Teilnehmer der Studie Stakeholder für die gezeigte Vision sein können. Um das zu erreichen, wurde ein Visionvideo gewählt, das die Funktionalitäten eines PKWs erweitert . Die Stakeholder einer solchen Vision wären hierbei die Autofahrer selber, Fußgänger, Radfahrer, und Nutzer der öffentlichen Verkehrsmittel. Der Grund für diese Menge an Stakeholdern ist, dass sowohl Nutzer als auch andere Teilnehmer im Verkehr, durch das Systems dieser Vision beeinflusst werden könnten. Da viele Menschen mindestens eines dieser vier Fortbewegungsmöglichkeiten nutzen, kommt ein Großteil der Bevölkerung Deutschlands als Stakeholder infrage.

Ein weiteres Kriterium war die Länge des Videos. Es sollte nicht zu kurz sein, um daraus genügend Anforderung extrahieren zu können, zeitgleich jedoch auch nicht zu lang, um die Aufmerksamkeit von den Stakeholdern über das gesamte Video lang aufrechtzuerhalten. Die Wahl fiel auf das Visionvideo „Future of Augmented reality (AR) in Cars“ [1], was die Funktionen eines Augmented Reality System eines Autos vorstellt, wie beispielhaft im Anhang A.1 zu sehen. Nachdem die Wahl getroffen wurde, konnten Anforderungen mithilfe des FunktionsMASTER-Models (siehe Kapitel 2.3) erstellt werden. Insgesamt wurden 27 Anforderungen erstellt, wobei 21 davon funktionale Anforderungen und 6 nicht-funktionale Anforderungen darstellen. Diese vorgefertigten Anforderungen dienen dazu, dass die Stakeholder keine Kenntnisse für die Erstellung von Anforderungen brauchen, sodass das Formulieren von neuen Anforderungen in der Umfrage komplett optional ist. Da sich diese Arbeit mit den Unterschieden von der Priorisierung nach Kano mit und ohne LLMs befasst, ist die Formulierung von Anforderungen nebensächlich und dient nur als Zwischenschritt.

Vor der Rekrutierung von potenziellen Stakeholdern wurde außerdem die Umfrage in Survey-SE erstellt. Zeitgleich der Um Umfrage wurden außerdem Handouts für die Priorisierung, eine Einverständniserklärung und standardisierte Formulare für beide Gruppen erstellt.

### 4.5.2 Umfrage

Die Umfrage wurde in Survey SE vor der Priorisierung (Schritt 2) durchgeführt und diente zum Erheben von Anforderung anhand des in der Vorbereitung erwähnten Visionvideos. Der erste Abschnitt befasste sich mit der Datenschutzerklärung (siehe Anhang A.2) und diversen demografischen Daten (siehe Anhang A.3, A.4, A.5). Anschließend sollten sich die Stakeholder das gesamte Visionvideo einmal in voller Länge anschauen, um eine Idee vom Produkt zu erhalten (siehe Anhang A.6). Das Video wurde für

[1] `https://www.youtube.com/watch?v=jPqN78-sbWo`

die nächsten Sektionen in acht Abschnitte unterteilt. In diesen konnten die Stakeholder für jede Anforderung entscheiden, ob sie dieser zustimmen oder nicht. Falls eine Anforderung nicht den Vorstellungen des Stakeholders entsprach, konnte diese per Textfeld umgeschrieben oder ganz gelöscht werden. Zusätzlich konnten Stakeholder für jeden Abschnitt eigene, neue Anforderungen aufschreiben (siehe Anhang A.7 - A.14).

### 4.5.3 Priorisierung

Nachdem die Umfrage ausgefüllt und die Einverständniserklärung beider Gruppen (Siehe Anhang A.18 - A.21) unterschrieben wurde, konnten Stakeholder ihre Anforderungen priorisieren. Die Priorisierung fand in Discord statt, was ein Onlinedienst für unter anderem Chat- und Sprachkonferenz ist [3]. Dort fanden sich der Moderator der Studie und Stakeholder zunächst in einem Sprachkanal zusammen, um den Vorgang erklärt zu bekommen. Zusätzlich standen Handouts zur Bedienung von ChatGPT (siehe Anhang A.16, A.17) und zum Kano-Modell (siehe Anhang A.15) zur Verfügung, die Stakeholder zur Hilfe nutzen konnten. Der Vorgang beider Gruppen verläuft leicht anders, hauptsächlich aus dem Grund, dass eine Gruppe ChatGPT nutzen kann. Zuerst werden also alle Teilnehmenden erinnert, zu welcher Gruppe sie gehören, da sich diese Information nur in der Einverständniserklärung befindet, die meistens schon am Vortag unterschrieben wurde. Anschließend erläutert der Moderator der Studie den Verlauf der Priorisierung beider Gruppen und betont dabei für die Stakeholder mit Zugang zu ChatGPT, dass sie ihre eigene Meinung teilweise oder komplett für den gesamten Vorgang der Priorisierung einfließen lassen können und so bei Bedarf ChatGPTs Entscheidungen ignorieren dürfen. Zusätzlich wird gebeten, ChatGPT-4o mini statt ChatGPT-4o zu verwenden, um Antworten stakeholderübergreifend so konsistent wie möglich zu halten. Nachdem der Verlauf erklärt und Fragen beantwortet wurden, haben Stakeholder per Privatnachricht ihr personalisiertes Formular ihrer jeweiligen Gruppe mit ihren Anforderungen erhalten, worin sie die Priorisierung vornehmen konnten. Anschließend wird das Startzeichen gegeben und die Timer für jeden Stakeholder gestartet, ohne diese davon in Kenntnis zu setzen. Jeder Stakeholder hat seinen eigenen Sprachkanal erhalten, damit bei Fragen die Anderen nicht gestört werden. Der Moderator der Studie stand bei Fragen im ursprünglich Sprachkanal zur Verfügung. Es durften keine inhaltlichen Fragen zu Kano, Priorisierung, ChatGPT oder ähnlichem an den Moderator gestellt werden. Lediglich Verständnisfragen wurden beantwortet. Im Gegensatz zu traditionellen Gruppe durften die anderen Stakeholder ChatGPT für jeweilige Unterstützung nutzen, unter anderem auch für die Beantwortung inhaltliche Fragen. Wurde das Formular vollständig ausgefüllt, schickte der Stakeholder ebenfalls per Privatnachricht das Dokument zurück und durfte anschließend den Server verlassen.

Die gesammelten Daten der Studie wurden anschließend mithilfe eines Spreadsheets aufbereitet und die nötigen Berechnungen und Grafiken erstellt.

### 4.5.4 Unterschiedlicher Priorisierungsverlauf der Gruppen

#### Verlauf mit ChatGPT

Zuerst öffnet der Stakeholder das Formular für die ChatGPT-Gruppe und ChatGPT mit einem beliebigen Browser. Solange noch unbearbeitete Anforderungen vorhanden sind, sollen diese Schrittweise priorisiert und die Begründung für die Priorisierung formuliert werden. Dieser Prozess kann optional durch ChatGPT unterstützt oder vollständig übernommen werden. Nach der Priorisierung einer Anforderung gibt der Stakeholder an, ob diese von ChatGPT übernommen oder eigenständig durchgeführt worden ist. Zum Schluss wird der kurze Frageteil, repräsentiert durch Likert-Skalen, durch ankreuzen beantwortet (siehe Anhang A.22, A.23), das Dokument abgespeichert und dem Studienleiter über Discord übermittelt. Die Fragen dienten dem Zweck, die Meinung von jedem Stakeholder bezüglich der Priorisierung mit ChatGPT als Unterstützungstool zu erfahren und damit die Forschungsfragen zu beantworten.

#### Verlauf ohne ChatGPT

Der Stakeholder öffnet das Formular für die traditionelle Gruppe und kann direkt mit der Priorisierung und Begründung jeder einzelnen Anforderung anfangen. Zum Schluss gibt es auch hier einen kurzen Frageteil zum ankreuzen (siehe Anhang A.24). Anschließend wird ebenfalls das Dokument abgespeichert und dem Studienleiter über Discord gesendet. Der Zweck der Fragen ist hier, genau wie bei der anderen Gruppe, die Erfahrungen mit Priorisierung in der Studie zu erhalten. Hier jedoch ohne die Unterstützung jeglicher Werkzeuge.

# Kapitel 5

# Ergebnisse

In diesem Kapitel werden die Ergebnisse der Studie und der Signifikantstests vorgestellt.

## 5.1 Demographie der Teilnehmer

Die Daten wurden von 16 TeilnehmerInnen zwischen 23 und 67 Jahren erhoben, wie Abbildung 5.1 veranschaulicht. Hier gehören die ersten acht Stakeholder zu der *ChatGPT-Gruppe* und demnach die letzten acht zur *traditionellen Gruppe*. Das Kürzel „S“ steht für studierend und „A“ für arbeitend, wobei Teilnehmer hier durch mangelnde Auswahlmöglichkeit unter A einmal „Arbeitslose“ und „Rentnerin“ angegeben haben, was hier jedoch keine negativen Folgen hat und „A“ als „nicht studierend, sondern ...“ angesehen werden kann. Die Zuteilung der Stakeholder in die Gruppen wird überwiegend zufällig durchgeführt. Einzig diejenigen, die keinen ChatGPT Account besitzen und ebenfalls keinen erstellen wollen, werden in die nicht-ChatGPT Gruppe eingeteilt.

| Nr | ID | Alter | Stud/Arb | Studiengang/Umfeld |
|---|---|---|---|---|
| 1 | 733652 | 29 | S | Technische Informatik |
| 2 | 189124 | 23 | S | Informatik |
| 3 | 130856 | 26 | A | Verwaltung |
| 4 | 160956 | 50 | A | Tierheim |
| 5 | 191732 | 29 | A | Einzelhandel |
| 6 | 199097 | 28 | A | Verkäufer |
| 7 | 823213 | 26 | S | Architektur |
| 8 | 931972 | 25 | S | Wirtschaftsinformatik |
| 9 | 73685 | 24 | S | Technische Informatik |
| 10 | 149447 | 67 | A | Rentnerin |
| 11 | 169373 | 26 | A | Gesundheitswesen - IT |
| 12 | 466334 | 24 | S | Informatik |
| 13 | 486458 | 24 | S | Informatik |
| 14 | 654351 | 25 | S | Informatik |
| 15 | 699245 | 24 | A | Gesundheitswesen - IT |
| 16 | 877835 | 42 | A | Arbeitslos |

Abbildung 5.1: Demografische Daten

## 5.2 Umfrage

In der Survey-SE Onlineumfrage wurden von den Stakeholdern unter anderem demografische Daten und Vorkenntnisse vor der Gruppenzuteilung erhoben, die hier vorgestellt werden.

Abbildung 5.2 zeigt den Median sowie das Mindest- und Maximalalter beider Gruppen in Jahren. Die Stakeholder mit ChatGPT sind um 2,5 Jahre älter, zählen jedoch auch den jüngsten Teilnehmer mit 23 Jahren zu ihrer Gruppe. Ebenfalls ist die älteste Teilnehmerin mit Zugang zu ChatGPT 17 Jahre jünger als die Älteste aus der traditionellen Gruppe.

### 5.2.1 Alter

| Gruppe | Median [Jahre] | Min [Jahre] | Max [Jahre] |
|---|---|---|---|
| mit GPT | 27 | 23 | 50 |
| ohne GPT | 24,5 | 24 | 67 |

Abbildung 5.2: Statistische Kennzahlen der Altersverteilung

### 5.2.2 Kenntnisse der Stakeholder

Um den Median der jeweiligen Vorkenntnisse zu berechnen und später in Kapitel 6 zu interpretieren, werden die Likert-Skalen hier folgendermaßen codiert:
1 : *Vertraut*
2 : *ein wenig vertraut*
3 : *kaum vertraut*
4 : *gar nicht vertraut*

Abbildung 5.3 zeigt die Vorkenntnisse beim Priorisieren von Anforderungen. Der Median der ChatGPT-Gruppe beträgt 3 und entspricht der Antwort *kaum vertraut*, der Median der traditionellen Gruppe 2 und ist gleichzusetzen zu *ein wenig vertraut*.

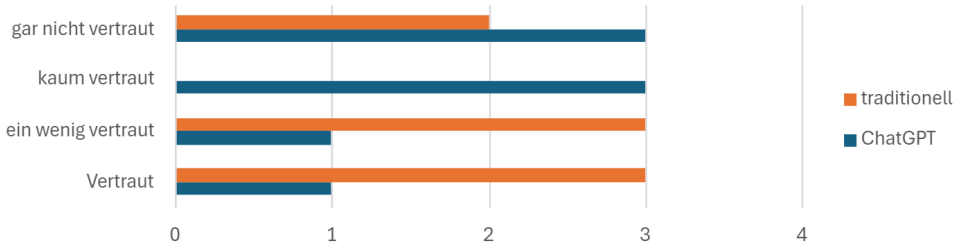



Abbildung 5.3: Vertrautheit mit Priorisierung

In Abbildung 5.4 sind die Vorkenntnisse der Stakeholder zum Kano-Modell zu sehen. Der Median der ChatGPT-Gruppe beträgt 4, der Median der traditionellen Gruppe 4.

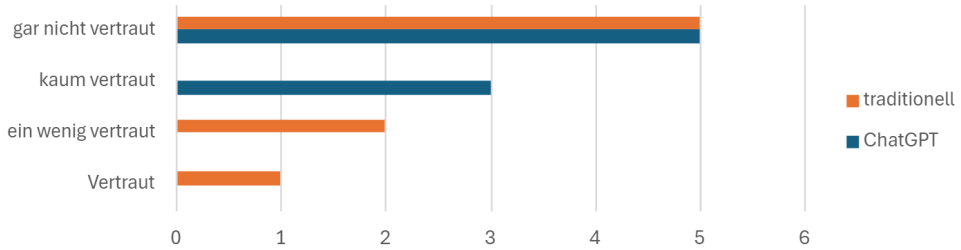



Abbildung 5.4: Vertrautheit mit Kano

In Abbildung 5.5 werden die Vorkenntnisse zu Large Language Models wie ChatGPT präsentiert. Der Median der ChatGPT-Gruppe beträgt 2, der Median der traditionellen Gruppe 1,5.

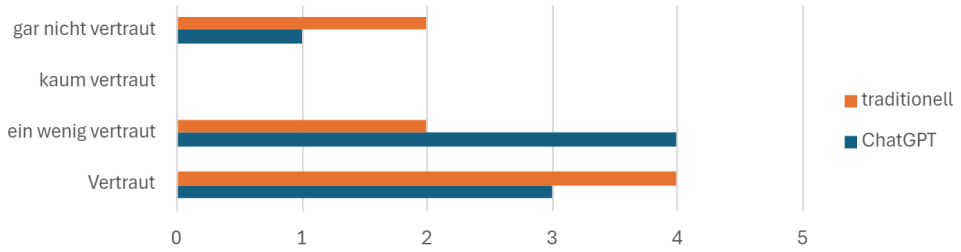



Abbildung 5.5: Vertrautheit mit Large Language Models

### 5.2.3 bevorzugte Fortbewegungsmittel

Für Abbildung 5.6 gaben beide Gruppen ihre bevorzugten Fortbewegungsmittel an, wobei mehrere Optionen ausgewählt werden konnten. Zur Auswahl standen die Fortbewegungsmittel, die im Visionvideo präsent waren.

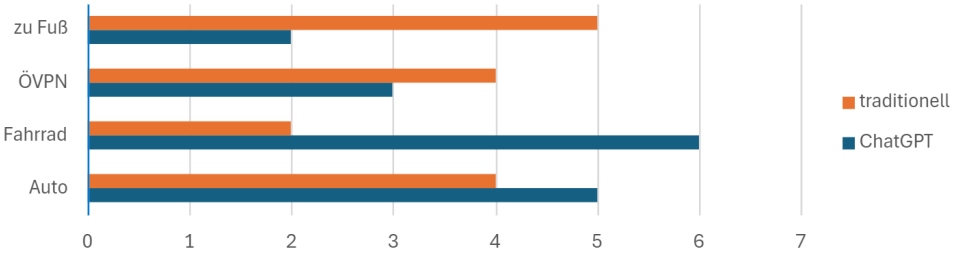



Abbildung 5.6: bevorzugte Fortbewegungsmittel

## 5.3 Priorisierung

Die folgenden Daten wurden in der letzten Phase der Studie, nach dem eigentlichen Priorisierungsprozess, gesammelt.

### 5.3.1 Anforderungen

Die Anforderungen der Stakeholder, inklusive der von ChatGPT übernommenen Priorisierungen, werden in Abbildung 5.7 tabellarisch dargestellt, wobei die ersten acht der ChatGPT-Gruppe angehören und die letzten acht der traditionellen Gruppe. Die Anzahl der Anforderungen unterscheidet sich, weil Stakeholder während der Onlineumfrage die vorgefertigten Anforderungen verändern, herausnehmen oder neue hinzufügen konnten. Zusätzlich zeigt die Abbildung in Spalte *Von ChatGPT übernommen* alle Anforderungen, deren Priorisierung von den Stakeholdern mithilfe von ChatGPT vorgenommen und übernommen wurden.

| Nr | Anforderungen gesamt | von ChatGPT übernommen | übernommener Anteil | Nr | Anforderungen gesamt |
|---|---|---|---|---|---|
| 1 | 30 | 30 | 100,00% | 9 | 26 |
| 2 | 29 | 21 | 72,41% | 10 | 27 |
| 3 | 27 | 25 | 92,59% | 11 | 27 |
| 4 | 27 | 26 | 96,30% | 12 | 28 |
| 5 | 25 | 23 | 92,00% | 13 | 23 |
| 6 | 27 | 26 | 96,30% | 14 | 27 |
| 7 | 27 | 25 | 92,59% | 15 | 27 |
| 8 | 27 | 26 | 96,30% | 16 | 27 |

Abbildung 5.7: Anforderungen der Stakeholder

Abbildung 5.8 zeigt die Summe aller Anforderungen beider Gruppen und den Gesamtanteil der von ChatGPT übernommenen Anforderungen mit 92,2%.

| Gruppe | Summe Anforderungen | von ChatGPT übernommen |
|---|---|---|
| mit GPT | 219 | 92,2% |
| ohne GPT | 212 | - |

Abbildung 5.8: Zusammenfassung der Anforderungen

### 5.3.2 Zeit

Im folgenden Unterkapitel werden die Statistiken zu den gemessenen Zeiten der Stakeholder während der Priorisierung präsentiert. Die Zeit jeden Stakeholders wurden mithilfe von Stoppuhren gemessen, wobei nur die Minuten und Sekunden betrachtet werden. Die Werte in Abbildung 5.9 beziehen sich auf die gemessenen Zeiten pro Anforderung jeder Gruppe.

| Gruppe | Median [Sekunden] | Min [Sekunden] | Max [Sekunden] |
|---|---|---|---|
| mit GPT | 93 | 71 | 116 |
| ohne GPT | 80 | 66 | 101 |

Abbildung 5.9: Statistische Kennzahlen der benötigten Zeit

Es folgen die nach dem Priorisierungsprozess erhobenen Daten. Um den Median der Likert-Skalen zu berechnen und später zu interpretieren, werden diese hier folgendermaßen codiert:
1 : *voll und ganz*
2 : *ein wenig*
3 : *neutral*
4 : *kaum*
5 : *gar nicht*

### 5.3.3 Frage 1

In Frage 1 haben die Stakeholder der ChatGPT-Gruppe beantwortet, in wie fern die Entscheidungen von ChatGPT, wie es Anforderungen nach Kano priorisiert hat, nachvollziehbar waren. Der Median der ChatGPT-Gruppe beträgt 2.

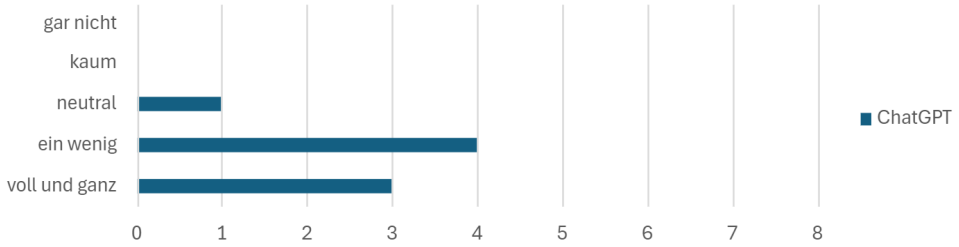



Abbildung 5.10: Wie sehr stimmen Sie mit der folgenden Aussage überein: die Entscheidungen von ChatGPT, wie es Anforderungen nach Kano priorisiert hat, waren nachvollziehbar.

### 5.3.4 Frage 2.1

Frage 2 beschäftigt sich mit dem Kano-Modell. In 2.1 gaben beide Gruppen an, zu welchem Grad das Kano-Modell verständlich war, wobei der Median der beider Gruppen 1 beträgt.

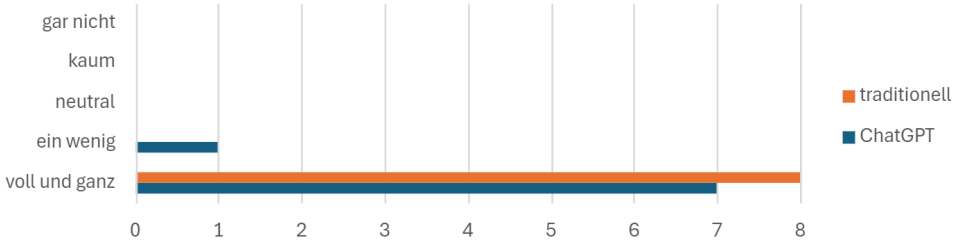



Abbildung 5.11: Wie sehr stimmen Sie mit der folgenden Aussage überein: das Kano-Modell war für mich verständlich.

### 5.3.5 Frage 2.2

Bei Frage 2.2 bewerteten die Stakeholder der ChatGPT-Gruppe, wie stark ihnen ChatGPT dabei geholfen hat, das Kano-Modell besser zu verstehen. Der Median dieser Bewertungen beträgt 5.

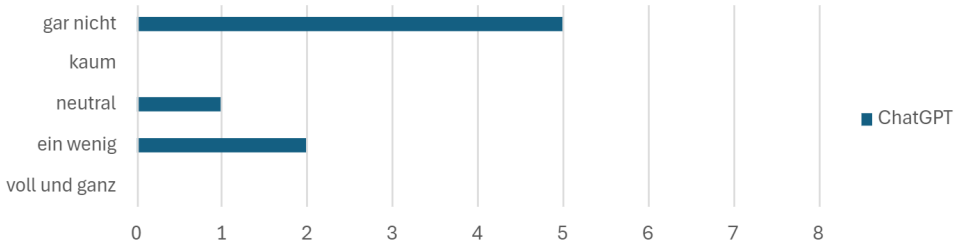



Abbildung 5.12: Wie sehr stimmen Sie mit der folgenden Aussage überein: ChatGPT konnte mir helfen, das Kano-Modell besser zu verstehen.

### 5.3.6 Frage 3.1

Frage 3 befasst sich mit der empfundenen Schwierigkeit beim Priorisieren. In 3.1 gaben beide Gruppen an, inwieweit es Momente gab, wo sie Schwierigkeiten beim Priorisieren von Anforderungen nach dem Kano-Modell hatten. Der Median der ChatGPT-Gruppe liegt bei 5, während der Median der traditionellen Gruppe bei 4 liegt.

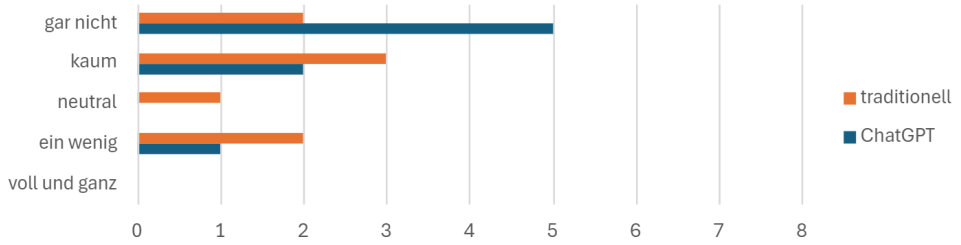



Abbildung 5.13: Wie sehr stimmen Sie mit der folgenden Aussage überein: es gab Momente, wo ich Schwierigkeiten hatte, eine Anforderung zu priorisieren.

### 5.3.7 Frage 3.2

Bei Frage 3.2 gab die ChatGPT-Gruppe an, zu welchem Grad ChatGPT beim Priorisieren von Anforderungen nach Kano eine Hilfe war, wobei der Median hier bei 1 liegt.

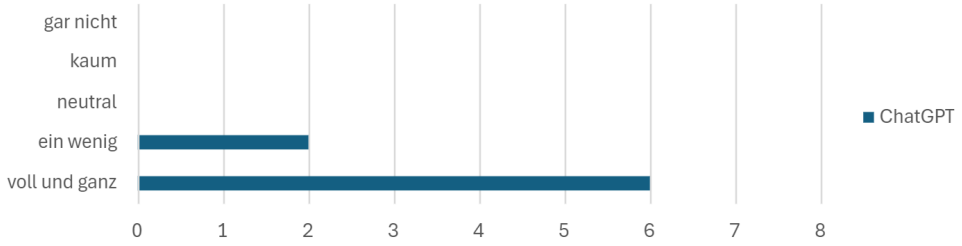



Abbildung 5.14: Wie sehr stimmen Sie mit der folgenden Aussage überein: ChatGPT war mir bei der Priorisierung von Anforderungen nach Kano eine Hilfe.

### 5.3.8 Frage 4.1

Die letzte Fragenkategorie, Frage 4, behandelt die Begründung der Priorisierung. In Frage 4.1 gaben beide Gruppen an, ob es Momente gab, in denen Schwierigkeiten bei der Begründung einer Priorisierung auftraten. Der Median der ChatGPT-Gruppe beträgt 5, der Median der traditionellen Gruppe 4.

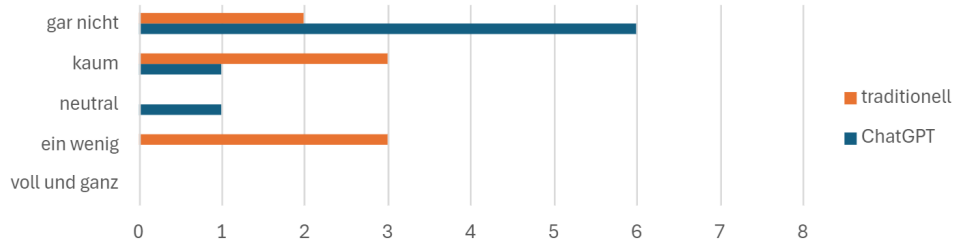



Abbildung 5.15: Wie sehr stimmen Sie mit der folgenden Aussage überein: es gab Momente, wo ich Schwierigkeiten hatten, eine Begründung für eine Priorisierung zu finden.

### 5.3.9 Frage 4.2

In der letzten Frage des Formulars wurden Stakeholder der ChatGPT Gruppe gefragt, in wie fern ChatGPT bei der Begründung für die Priorisierung helfen konnte. Der Median bei dieser Frage beträgt 1.

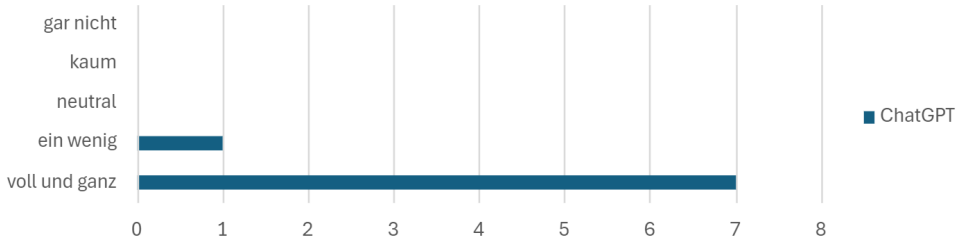



Abbildung 5.16: Wie sehr stimmen Sie mit der folgenden Aussage überein: ChatGPT konnte mir bei der Begründung der Priorisierungen helfen.

### 5.3.10 Signifikantstest

Tabelle 5.1 zeigt die Werte der Signifikantstests für die in Kapitel 4.2 aufgestellten Hypothesen der Studie. Die Ergebnisse des Mann-Whitney-U-Test wurden mithilfe der Funktion *mannwhitneyu* aus Pythons *scipy.stats*-Modul berechnet.

| **Hypothese** | **U-Wert** | **Z-Wert** | **p-Wert** |
|---|---|---|---|
| H1 | 46,5 | 1,5 | 0,1411 |
| $H2_1$ | 45 | 1,37 | 0,1623 |
| $H2_2$ | 49,5 | 1,84 | 0,0536 |
| $H_{Kano}$ | 36 | 0,42 | 0,3816 |

Tabelle 5.1: Ergebnisse des Mann-Whitney-U-Test

# Kapitel 6

# Diskussion

In diesem Kapitel werden die Forschungsfragen beantwortet und die Ergebnisse aus Kapitel 5 interpretiert. Zur besseren Übersicht werden dabei die Abkürzungen $G_{ai}$ für die ChatGPT-Gruppe und $G_t$ für die traditionelle Gruppe verwendet.

## 6.1 Beantwortung der Forschungsfragen

In diesem Abschnitt werden die in Kapitel 1.2 aufgestellten Forschungsfragen basierend auf den erarbeiteten Ergebnissen beantwortet.

**[RQ1]**: *Wie unterscheidet sich die Geschwindigkeit der Priorisierung von Anforderungen mit LLMs in Vergleich zu ohne?*

Forschungsfrage eins wird anhand der interpretierten Ergebnisse der gemessenen Zeit beantwortet. Die Analyse der benötigten Zeit beim Priorisierungsprozess zeigte, dass Stakeholder mit ChatGPT mehr Zeit benötigten, als diejenigen, die traditionelle Methoden ohne LLMs verwendeten. Das bedeutet, dass, entgegen den Erwartungen, ChatGPT den Prozess insgesamt verlangsamt. Diese Beobachtung könnte durch Faktoren wie mangelnde Vorkenntnisse, langes Warten auf ChatGPT's Antworten oder langsamer Internetverbindung entstanden sein, wobei der Unterschied der benötigten Zeit beider Gruppen statistisch nicht signifikant ist und damit zufällig entstanden sein könnte.

**[RQ2]**: *Wie unterscheidet sich die Schwierigkeit der Anforderungspriorisierung mit LLMs in Vergleich zu ohne?*

Die zweite Forschungsfrage wird mithilfe den Ergebnissen aus $NH2_1$, $NH2_2$, Frage 3.1 und 4.1 beantwortet. Der Priorisierungsprozess umfasste sowohl die Einordnung der Anforderungen in die Kano-Kategorien als auch die Begründung dieser Entscheidungen. Beide Schritte, sowohl das Priorisieren, als auch das Begründen, wurden von den Stakeholdern, die

Zugang zu ChatGPT hatten, als etwas leichter empfunden als von der Gruppe, die auf traditionelle Methoden angewiesen war. Auch wenn keine statistisch signifikanten Unterschiede in der empfundenen Schwierigkeit zwischen den beiden Gruppen nachgewiesen werden konnten, bleibt die Beobachtung dennoch interessant, dass ChatGPT den Prozess für die Stakeholder etwas erleichtern kann.

**[RQ3]**: *Kann ChatGPT Stakeholder beim Priorisieren von Anforderungen unterstützen?*

Um die dritte Forschungsfrage zu beantworten, werden hauptsächlich die Ergebnisse der Frage 2.2, 3.2 und 4.2 herangezogen. Die Ergebnisse der Studie zeigen, dass ChatGPT überwiegend eine unterstützende Rolle im Priorisierungsprozess einnimmt. Die Mehrheit der Stakeholder empfand ChatGPT als wertvolle Hilfe sowohl beim Priorisieren als auch beim Begründen der Priorisierungen. Diese Beobachtung wird durch die hohe Übernahmerate der von ChatGPT vorgeschlagenen Priorisierungen und die solide Nachvollziehbarkeit der Antworten weiter gestützt. Einzig beim Verständnis des Kano-Modells waren sich die meisten Stakeholder einig, dass ChatGPT nicht besonders hilfreich war. Dies könnte jedoch darauf zurückzuführen sein, dass die Erklärungen und die bereitgestellten Handouts in der Studie ausreichendes Grundverständnis vermittelten, sodass der Einsatz von ChatGPT in diesem Bereich nicht mehr notwendig war. Zusammenfassend lässt sich sagen, dass ChatGPT effektiv den Stakeholder beim Priorisieren von Anforderung unterstützen kann.

## 6.2 Interpretation der Ergebnisse

In diesem Abschnitt werden die erhobenen Daten zunächst kurz zusammengefasst und anschließend interpretiert. Zunächst widmen wir uns den Ergebnissen der Onlineumfrage sowie der Priorisierungsstatistik aus 5.3. Anschließend behandeln wir die Fragen, die nur Stakeholder der ChatGPT-Gruppe beantworteten und schließen das Unterkapitel mit dem Überprüfen der aufgestellten Hypothesen.

### 6.2.1 Ergebnis Alter

Der Median, also das mittlere Alter der Stakeholder, ist bei $G_{ai}$ mit 27 um 2,5 Jahre größer und deutet darauf hin, dass $G_{ai}$ etwas älter ist als $G_t$. Auffälig ist noch, dass der jüngste Stakeholder in $G_{ai}$ mit 23 Jahren nur ein Jahr jünger ist als der Jüngste in $G_t$. Allerdings beträgt der Altersunterschied zwischen den beiden ältesten Stakeholdern 17 Jahre, wie 5.2 zeigt. Ein Altersunterschied zwischen den Gruppen könnte sich auf verschiedene Weisen auf die Ergebnisse der Studien auswirken. Jüngere Menschen haben tendenziell mehr Erfahrung beim Umgang mit Computern

und können mit dem Wissen und einer schnelleren Tippgeschwindigkeit auf der Tastatur die Aufgaben schneller erledigen, was sich direkt auf die gemessene Zeit auswirken würde. Um solchen Einfluss zu minimieren, wurden gezielt die beiden ältesten Teilnehmerinnen in verschiedene Gruppen zugeordnet. Der Altersunterschied ist zwar vorhanden, beträgt jedoch nur 2,5 Jahre und ist daher nicht so erheblich, dass signifikante Unterschiede in den Ergebnissen allein auf das Alter zurückzuführen sind. Der p-Wert des Alters von 0,3951 liegt allerdings deutlich über dem Signifikanzniveau von 0,05, was darauf hinweist, dass der Unterschied im Alter zwischen den beiden Gruppen zwar vorhanden, aber statistisch nicht signifikant ist.

### 6.2.2 Ergebnis Kenntnisse

Die Vorkenntnisse der Stakeholder wurden mit einer 4-stufigen Likert-Skala erhoben und die Ergebnisse hier interpretiert.

Vertrautheit von Priorisierung: Der Median von $G_{ai}$ ist 3, was einem *kaum vertraut* entspricht, während der Median von $G_t$ 2 beträgt, also *ein wenig vertraut*. Der Unterschied im Median zwischen beiden Gruppen deutet darauf hin, dass $G_t$ ein etwas höheres Maß an Erfahrung mit Priorisierung aufweist. Die Folge dieses Unterschieds könnte bewirken, dass die unabhängigen Variablen dieser Arbeit dadurch beeinflusst werden, was in einem optimalen Szenario nicht vorkommen sollte.

Vertrautheit vom Kano-Modell: Der Median von beiden Gruppen beträgt 4, was bedeutet, dass die Mehrheit der Stakeholder *gar nicht vertraut* mit dem Kano-Modell sind. Auch wenn drei Stakeholder aus $G_t$ mindestens ein wenig vertraut mit dem Thema sind, so ist der Wissensstand hierbei deutlich näher als bei der Priorisierung, was wünschenswert ist, um die Ergebnisse so wenig wie möglich vom Kenntnisstand abhängig zu machen.

Vertrautheit von Large Language Models: Hier wurden alle 16 Stakeholder befragt, weil die Gruppenzuweisung erst nach der Umfrage stattfand. Demnach sind die Antworten der Stakeholder aus $G_t$ im Grunde nicht von Interesse. Die Mediane verdeutlichen allerdings, dass beide Gruppen relativ ausgeglichen im Kenntnisstand von LLMs sind und kein Stakeholder bei der Gruppenfindung bevorzugt behandelt wurde. Der Median von $G_{ai}$ ist 2, was einem *ein wenig vertraut* entspricht und bedeutet, dass überwiegend ein grundlegendes Verständnis bei LLMs wie ChatGPT vorliegt. Das ist beim Priorisierungsprozess natürlich vom Vorteil, weil sich so die Stakeholder besser mit ChatGPT einfinden konnten.

### 6.2.3 Ergebnis bevorzugtes Fortbewegungsmittel

Es standen vier Fortbewegungsmittel zur Verfügung, die Stakeholder per Mehrfachauswahl wählen konnten. Die Daten könnten benutzt werden, um eine Tendenz von der Wahl der Priorisierung in eine Kano-Kategorie zu finden. So könnten Stakeholder, die zu Fuß gehen, eine Anforderung als weniger wichtig erachten, als Stakeholder, die das Auto bevorzugen. Da diese Daten für die Forschungsfragen allerdings nicht direkt eine Rolle spielen, dienen diese Daten nur zur Veranschaulichung, dass die TeilnehmerInnen dieser Studie tatsächlich Stakeholder dieses Produkts sind.

### 6.2.4 Ergebnis Anforderungen

Die Abbildung 5.7 zeigt die Anforderungen beider Gruppen, wobei sich die Anzahl teilweise unterscheidet. Daher wird auch nicht die Gesamtzeit, sondern die Zeit pro Anforderung für diese Arbeit betrachtet. Abbildung 5.8 zeigt eine durchschnittliche Übernahmerate von 92,2%, was schlussfolgernd bedeutet, dass die meisten Entscheidungen von ChatGPT bezüglich der Priorisierung in die Kano-Kategorien mit der Meinung der Stakeholder übereinstimmt, oder dass die Begründung für diese Entscheidungen überzeugend waren. Eine so hohe Übernahmerate spricht dafür, dass ChatGPT die Entscheidung der bevorzugten Priorisierung erleichtern, oder sogar ganz übernehmen kann.

### 6.2.5 Ergebnis Nachvollziehbarkeit von GPTs Antworten (Frage 1)

Die Verteilung der Antworten in Abbildung 5.10 zeigt, dass die Mehrheit der Stakeholder die Entscheidungen von ChatGPT als gut nachvollziehbar empfanden. Mit einem Median von 2 und der Tatsache, dass sieben von acht Stakeholder eine gute bis sehr gute Bewertung abgegeben habe, kann angenommen werden, dass ChatGPTs Antworten überwiegend verständlich waren. Diese Ergebnisse stimmen mit der hohen Übernahmerate aus Abbildung 5.8 überein.

### 6.2.6 Ergebnis GPTs Hilfe, Kano besser zu verstehen (Frage 2.2)

Obwohl das Kano-Modell von den Stakeholdern überwiegend verstanden worden ist, zeigt das Ergebnis der Likert-Skala in Abbildung 5.11, dass ChatGPT größtenteils nicht zu diesem Verständnis beigetragen hat. Fünf der acht Stakeholder hat ChatGPT in dieser Hinsicht gar nicht geholfen, wodurch auch der Median von 5 resultiert. Das Ergebnis kann so interpretiert werden, dass ChatGPT's Erklärungen zum Kano-Modell nicht verständlich waren. Wir denken jedoch, dass, wie in 6.2.9 schon erwähnt, das Handout

ausreichendes Grundverständnis vermittelte und ChatGPT somit gar nicht befragt wurde. Andererseits gibt es auch zwei Fälle, wo ChatGPT *ein wenig* geholfen hat. Hier haben Stakeholder ChatGPT zusätzlich zur Erklärung des Kano-Modells herangezogen und sind der Meinung, dass die Antworten gut, aber nicht perfekt waren. Wenn also tatsächlich kein Erklärungsbedarf herrscht, ist ChatGPT ein gutes Werkzeug, kann aber bei komplexeren Themen eventuell keinen Experten ersetzen, der sein Fachwissen den Stakeholdern verständlich übermitteln könnte.

### 6.2.7 Ergebnis GPTs Hilfe bei Priorisierung (Frage 3.2)

Abbildung 5.14 zeigt die Antworten der Stakeholder aus $G_{ai}$ bezüglich der Unterstützung durch ChatGPT bei der Priorisierung. Von den 8 Stakeholdern gaben 6 an, dass sie der Aussage *voll und ganz* zustimmen, während 2 *ein wenig* zustimmten, was zu einem Median von 1 führt. Dieses Ergebnis deutet darauf hin, dass die Mehrheit der Stakeholder aus $G_{ai}$ die Unterstützung von ChatGPT bei der Priorisierung von Anforderungen als nützlich empfand. Selbst jene, die nicht uneingeschränkt zustimmten, sahen zumindest einen gewissen Nutzen. Daraus lässt sich ableiten, dass ChatGPT eine unterstützende Funktion im Bereich der Priorisierung nach Kano bietet und den Stakeholdern bei diesem Prozess entlastend zur Seite stehen kann.

### 6.2.8 Ergebnis GPTs Hilfe bei Begründung (Frage 4.2)

In Abbildung 5.16 wurden Stakeholder aus $G_{ai}$ gefragt, wie sehr ihnen ChatGPT beim Begründen ihrer Priorisierungen helfen konnte. Sieben Stakeholder gaben die beste Wertung mit *voll und ganz* an, ein Stakeholder die zweitbeste Bewertung *ein wenig*. Dementsprechend entspricht der Median einer 1. Diese Verteilung verdeutlicht, dass die Mehrheit der Stakeholder die Unterstützung von ChatGPT bei der Begründung als äußerst hilfreich empfanden. Zwar geben die Studienergebnisse nicht explizit Aufschluss darüber, ob die Begründungen vollständig übernommen oder teilweise durch eigenen Text ergänzt wurden, doch ein deutlicher Unterstützungsfaktor ist in jedem Fall vorhanden. Insbesondere bei einer Vielzahl von Anforderungen dieser komplexen und zeitaufwändigen Aufgabe kann ChatGPT eine Hilfestellung bieten und die Findung von Argumenten erleichtern. Die Ergebnisse legen also nah, dass ChatGPT ein effektives Werkzeug für Stakeholder darstellt, um die Formulierung der Begründung zu erleichtern.

### 6.2.9 Signifikanz der Unterschiede zwischen den Gruppen

Die Ergebnisse der Mann-Whitney-U-Tests werden in diesem Abschnitt interpretiert, indem zunächst die Hypothese wiederholt und anschließend die relevanten Daten analysiert werden. Vorab ist anzumerken, dass eine statistische Signifikanz ab einem p-Wert von unter 5% angenommen wird.

Das bedeutet, dass eine Hypothese nur dann akzeptiert werden kann, wenn die Nullhypothese mit einem p-Wert von unter 0,05 verworfen wird. Der p-Wert gibt dabei die Wahrscheinlichkeit an, die beobachteten Daten zu erhalten, wenn die Nullhypothese zutrifft. Beispielsweise bedeutet ein p-Wert von 0,15, dass es eine 15%ige Wahrscheinlichkeit gibt, die beobachteten Unterschiede zu sehen, falls die Nullhypothese wahr ist. Ein p-Wert von 0,15 würde daher keine ausreichenden Beweise liefern, um die Nullhypothese zu verwerfen, was darauf hindeutet, dass kein signifikanter Unterschied zwischen den Gruppen besteht.

### Ergebnis Zeit

Das Ergebnis dieses MWU-Tests untersucht die Hypothese $H1$, die lautet: „Es gibt einen signifikanten Unterschied zwischen der für die Priorisierung durchschnittlich benötigten Zeit pro Anforderung beider Gruppen."

Abhängig vom Ergebnis kann hier gedeutet werden, ob die Nutzung von ChatGPT beim Priorisieren von Anforderungen zur Effizienz, also zum einsparen von Zeit, beiträgt. Dafür wurden für jeden Stakeholder die benötigte Zeit für das Priorisieren für jede Anforderung betrachtet und mit diesen Daten der MWU durchgeführt. Das Ergebnis zeigt, dass mit einem p-Wert von 0,1411 die Nullhypothese $NH1$ nicht verworfen werden kann und es keinen signifikanten Unterschied zwischen den Zeiten beider Gruppen besteht. Die Abbildung 5.9 zeigt, dass der Median der Zeit für das Priorisieren bei $G_{ai}$ um 13 Sekunden pro Anforderung höher ist. Das bedeutet, dass Stakeholder mit ChatGPT länger benötigen, was zunächst verwunderlich erscheint, sich aber erklären lässt. Ein Faktor könnte natürlich das Alter der Stakeholder sein, weil ältere Menschen dazu neigen, Computer und vor allem neuartige Software langsamer zu bedienen. Wie allerdings schon festgestellt, ist die Chance sehr hoch, dass das Alter zwischen den Gruppen keinen signifikanten Unterschied aufzeigt. Auch der Median mit nur 2,5 Jahre mehr bei $G_{ai}$ zeigt, dass das Alter der Gruppen balanciert ist. Ein weiterer Faktor ist die Vorkenntnis der Stakeholder bezüglich der Priorisierung von Anforderung. Der Median von 3 (kaum vertraut) bei $G_{ai}$ und einem Median von 2 (ein wenig vertraut) bei $G_t$ zeigt, dass die Stakeholder mit ChatGPT weniger vertraut mit der Priorisierung sind und so mehr Zeit für die Einarbeitung benötigen als die traditionelle Gruppe. Ein Einfluss kann jedoch auch auf andere Quellen zurückzuführen sein. So konnte nach der Studie beobachtet werden, dass die Begründung, die zum Anforderungsprozess dazuzählt, bei $G_{ai}$ teilweise deutlich länger ist als bei $G_t$. Falls der Stakeholder sich nicht entscheidet, ChatGPT im Prompt anzuweisen, die Begründung kurz zu halten, wurde meistens ein ganzer Absatz generiert. Dieser Prozess, inklusive dem Kopieren und Einfügen in das Dokument, benötigt Zeit. Um diese Vermutung zu bestätigen, führte der Moderator nach der Studie den

Priorisierungsprozess anhand der Basisanforderungen nochmal aus und konnte beobachten, dass die Generierung der Antwort pro Anforderung bis zu 5 Sekunden in Anspruch nehmen kann. Dieser Wert variiert zwar je nach Hardware, Internetverbindung und Prompt, zeigt aber, dass ein signifikanter Anteil der Zeit auf das Warten auf ChatGPT zurückzuführen ist. Falls sich Stakeholder entscheiden, die Antworten via Prompt oder Regenerate-Funktion abzuändern, wäre der Prozess noch länger.

Abschließend lässt sich sagen, dass ChatGPT keine Alternative zur traditionellen Methode ist, wenn es um Zeitersparnis geht

### Ergebnis Kano Verständlichkeit (Frage 2.1)

Das Ergebnis dieses MWU-Tests untersucht die Hypothese $H_{Kano}$, die lautet: „Es gibt einen signifikanten Unterschied zwischen dem Verständnis beider Gruppen bezüglich des Kano-Modells."

Ein p-Wert von 0,3816 liegt deutlich über dem Signifikanzniveau von 0,05, was bedeutet, dass die Nullhypothese $NH_{Kano}$ nicht verworfen werden kann. Es gibt also keinen signifikanten Unterschied zwischen dem Verständnis beider Gruppen bezüglich des Kano-Modells. Zudem gaben 15 Stakeholder an, das Kano-Modell voll und ganz verstanden zu haben, wie in Abbildung 5.11 zu sehen. Diese Ergebnisse deutet darauf hin, dass die Nutzung von ChatGPT das Verständnis des Kano-Modells nicht maßgeblich beeinflusst hat. Besonders auffällig ist, dass 13 der 16 Stakeholder vor der Studie kaum bis gar nicht mit dem Kano-Modell vertraut waren, in Abbildung 5.14 aber später angaben, es voll und ganz verstanden zu haben. Daher lässt sich mit den vorliegenden Ergebnissen vermuten, dass traditionelle Methoden wie eine mündliche Erklärung und ergänzende Handouts ausreichend waren, um das notwendige Verständnis zu vermitteln. ChatGPT war in diesem Kontext nur vereinzelt von Nutzen, um Verständnislücken zu füllen.

### Ergebnis Schwierigkeiten bei Priorisierung (Frage 3.1)

Das Ergebnis dieses MWU-Tests untersucht die Hypothese $H2_1$ , die lautet: „Es gibt einen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Priorisieren von Anforderungen."

Der p-Wert des Mann-Whitney-U-Test für die Frage 3.1 (siehe Abbildung 5.13) ergibt einen p-Wert von 0,1623, der über dem Signifikanzniveau von 0,05 liegt. Das bedeutet, dass die Nullhypothese $NH2_1$ nicht verworfen werden kann. Somit gibt es keinen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Priorisieren von Anforderungen. Die Auswertung zeigt außerdem einen Median von 5 bei $G_{ai}$ und 4 bei $G_t$. Das deutet darauf hin, dass Stakeholder mit ChatGPT überwiegend gar keine Schwierigkeiten beim Priorisieren empfanden, während diejenigen ohne ChatGPT tendenziell etwas mehr Probleme hatten. In Kombination mit den

Ergebnissen aus Abschnitt 6.2.7 könnte eine Korrelation vermutet werden, dass ChatGPT für diesen Unterschied verantwortlich ist. Auch wenn die Nullhypothese also nicht abgelehnt werden kann, ist es dennoch interessant zu beobachten, dass ChatGPT möglicherweise einen positiven Einfluss auf den Priorisierungsprozess hat. Um diesen potenziellen Unterschied statistisch signifikant zu machen, könnte eine zukünftige Studie mit einer größeren Teilnehmerzahl durchgeführt werden, um zu zeigen, dass ChatGPT tatsächlich den Priorisierungsprozess erleichtern kann.

**Ergebnis Schwierigkeit bei Begründung (Frage 4.1)**

Das Ergebnis dieses MWU-Tests untersucht die Hypothese $H2_2$ , die lautet: „Es gibt einen signifikanten Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Formulieren einer Begründung für die Priorisierung.“

Der p-Wert von 0,0536 liegt knapp über dem festgelegten Signifikanzniveau von 0,05, sodass die Nullhypothese auch in diesem Fall nicht verworfen werden kann. Dies bedeutet, dass kein signifikanter Unterschied zwischen der empfundenen Schwierigkeit beider Gruppen beim Formulieren einer Begründung für die Priorisierung festgestellt werden kann. Da der Wert jedoch sehr nah an der Schwelle zur Signifikanz liegt, lohnt es sich hier, die Antworten der Stakeholder vorzustellen und genauer zu betrachten. Die Analyse der Abbildung 5.15 zeigt deutliche Unterschiede zwischen den beiden Gruppen. Der Median der Gruppe $G_{ai}$ liegt bei 5, was darauf hinweist, dass die meisten Stakeholder dieser Gruppe *gar nicht* auf Schwierigkeiten gestoßen sind. $G_t$ hingegen hatte vermutlich mehr Schwierigkeiten bei der Begründung, wie auch der Median von 4 vermuten lässt. Besonders auffällig ist, dass 6 der 8 Stakeholder mit Zugriff auf ChatGPT überhaupt keine Schwierigkeiten hatten, während in der anderen Gruppe nur 2 Stakeholder dieser Meinung waren. Auf der anderen Hälfte der Skala gaben ausschließlich 3 Stakeholder aus $G_t$ an, *ein wenig* Schwierigkeiten gehabt zu haben. Diese Ergebnisse lassen, in Kombination mit den Erkenntnissen aus Abschnitt 6.2.8, auf eine Korrelation zwischen der Nutzung von ChatGPT und dem unterschiedlich empfundenen Schwierigkeitsgrad schließen. Dort wurde festgestellt, dass ChatGPT als sehr hilfreich beim Formulieren von Begründungen empfunden wurde. Dieser Unterschied zwischen beiden Gruppen könnte daher tatsächlich auf die Nutzung von ChatGPT zurückzuführen sein, wobei der Unterstützungsfaktor noch deutlicher wird, wenn man bedenkt, dass die Vorkenntnisse der Stakeholder ohne ChatGPT etwas höher sind (siehe Abschnitt 6.2.2) und so eigentlich erwartet werden könnte, dass diese auch weniger Schwierigkeiten beim Prozess hätten. Zusammenfassend lässt sich sagen, dass die Ergebnisse für die Nutzung für ChatGPT beim Formulieren der Begründung für die Priorisierung sprechen. Auch wenn die Nullhypothese nicht verworfen werden

kann, liegt der p-Wert sehr nahe an der 5%-Schwelle, weshalb auch hier eine Studie mit deutlich mehr Stakeholdern zusätzliche Einblicke ermöglichen könnte.

## 6.3 Threats to Validity

Um die Einschränkungen der Studienergebnisse zu verdeutlichen, werden hier nach Wohlin et al. [29] internal, external, conclusion und construct validity besprochen.

### Internal Validitiy

Internal Validity beschreibt, ob die untersuchten Effekte der Studie auf die unabhängigen Variablen zurückzuführen sind. Verletzt werden kann diese validity durch äußere Einflussfaktoren, wie zum Beispiel von den Teilnehmern unterschiedlich genutzte Hardware für die Studienbearbeitung.

Da diese Studie komplett online durchgeführt wird, kann nicht garantiert werden, dass Stakeholder unerlaubte Tools eingesetzt haben oder sich während der Bearbeitung ablenkten. Auch kann nicht garantiert werden, dass die besprochene ChatGPT-Version benutzt wurde, da selbst eine Bildschirmübertragung und konstante Überwachung nicht garantieren, dass Stakeholder mit einem anderen Gerät diese Tools nutzen. Außerdem kann nicht garantiert werden, dass Stakeholder sich durch das Formular klicken. Es wurden jedoch immer alle Begründungen angegeben, was zeigt, dass Stakeholder die Aufgaben wie vorgesehen bearbeitet haben. Eine Offlinestudie hätte diese Gefahren wahrscheinlich eliminieren können, würde aber auch zu einer kleineren Teilnehmeranzahl führen, da bei einer erhebliche Anzahl an Probanden nicht erwarten werden konnte, dass diese zur Leibniz Universität Hannover kommen. Bezüglich der Zeit kann nicht garantiert werden, dass jeder das gleiche Setup hat, weshalb internet- oder hardwarebedingt Stakeholder mal schneller, mal langsamer die Aufgabe erledigen könnten. Das sollten jedoch sehr kleine Unterschiede sein, da ein offener Browser mit ChatGPT vergleichbar geringe Ressourcen benötigt. Eine weitere Sache wären Fähigkeiten der Stakeholder, die nicht abgefragt wurden. Z.B. können schnelle Schreiber die Aufgabe schneller erledigen. Allgemein können Leute mit EDV-Wissen, wie das Nutzen von Keyboard shortcuts. In einer realen Umgebung würden Stakeholder diese Unterschiede allerdings auch mitbringen. Technische Ausfälle der Hardware, Discord oder ChatGPT waren stets ein Risiko, wobei sich jedoch keine der Stakeholder beschwerte. Die Verfügbarkeit von ChatGPT und Discord kann hier jedoch durch den Durchführenden der Studie bestätigt werden. ChatGPT lernt mit jedem Prompt und basiert seine Antwort mit Bezug auf den Chatverlauf, weshalb die Konsistenz der Antworten über alle Anforderungen hinweg gefährdet sein könnte. Anhand der Begründungen von der ChatGPT-Gruppe

konnte allerdings eine überwiegend konsistente Form und Länge festgestellt werden, was auf konsistente Antworten seitens ChatGPT hinweisen kann. Zwischen den Stakeholdern gab es aber natürlich unterschiedliche pattern der Antworten, was wahrscheinlich auf verschiedene Angehensweisen des prompt engineering der Stakeholder zurückzuführen ist. Mit pattern ist hier gemeint, dass sich überwiegend eine gewisse gemeinsame Struktur und Länge der Antworten ergibt. Das wurde jedoch erwartet, da es kein Zwang war, die Beispielprompts aus dem Handout zu übernehmen. Zum Schluss sei gesagt, dass die Studie zu verschiedenen Zeiten an verschiedenen Wochentagen durchgeführt, was Laune und Motivation beeinflussen könnte.

### External Validity

External Validity beschreibt die Generalisierbarkeit von Ergebnissen. Hierbei können externe Faktoren wie Laune der Teilnehmenden oder das Wetter Ergebnisse generalisierbar machen. Faktoren, die die External Validity positiv beeinflussen, wirken sich negativ auf die Internal Validity aus.

In dieser Studie bestehen Teilnehmende hauptsächlich aus Studenten und Arbeitnehmer, die auch als Stakeholder des Produkts angesehen werden können, wobei diese per Convenience Sampling rekrutiert worden sind. Durch diese nicht-zufällige Stichprobenmethode, bei der Teilnehmer aufgrund ihrer einfachen Erreichbarkeit ausgewählt werden, könnte die External Validity gefährdet sein, da die Stichprobe möglicherweise nicht die gesamte Zielpopulation repräsentiert. Überwiegend sind es junge Erwachsene unter 30, aber auch drei zwischen 40 und 70, was zur Generalisierung der Ergebnisse beiträgt. Problem bei den Studierenden ist, dass sie hauptsächlich an der Leibniz Universität Hannover studieren und so die Generalisierbarkeit für andere Studenten an anderen Universitäten bedroht ist. Alle Stakeholder waren außerdem ausschließlich in Deutschland lebende Erwachsene, wodurch das System auf andere Länder wohlmöglich nicht generalisierbar ist. Dass Stakeholder ihre eigenen Geräte für die Priorisierung nutzen dürfen, trägt zur Generalisierbarkeit bei. In einem CrowdRE-setting kann man erwarten, dass viele verschiedene Stakeholder ihr bevorzugtes Gerät für den Vorgang nutzen.

### Conclusion Validity

Die Conclusion Validity beschreibt die Korrektheit von Schlussfolgerungen aus einer Studie. Diese kann unter anderem durch eine zu kleine Stichprobe an Teilnehmenden gefährdet sein. Eine Gesamtmenge von 16 TeilnehmerInnen könnte zu gering sein, um eine gute Stichprobe möglicher Stakeholder des Produkts im Visionvideo vorzuweisen. Anfangs waren für die Studie 20 Stakeholder eingeplant, was trotz Bemühungen nicht erreicht werden konnte. Die auftretenden Probleme könnten unter anderem sein, dass starke

Ausreißer großen Einfluss auf die Ergebnisse haben, die bei einer größeren Stichprobe weniger ins Gewicht fallen würden. Es führt außerdem dazu, dass bei einer Anzahl bei acht Stakeholdern pro Gruppe, eine statistische Signifikanz schwerer zu erreichen ist.

### Construct Validity

Die Construct Validity besagt, inwiefern die Studie das misst, was wir tatsächlich herausfinden möchten. In der Studie wird unter anderem die subjektiv wahrgenommene Schwierigkeit mit einer 5-stufigen Likert-Skala abgefragt. Bei einer subjektiven Erfassung können viele äußere Faktoren das Ergebnis dieser Befragung verändern, weshalb dahingehend die Construct Validity beeinflusst werden könnte.

# Kapitel 7

# Zusammenfassung und Ausblick

Abschließend wird in diesem Kapitel eine Zusammenfassung der Arbeit und ein Ausblick gegeben.

## 7.1 Zusammenfassung

Das Ziel dieser Arbeit war es zu untersuchen, ob das Large Language Model *ChatGPT* den Prozess der Anforderungspriorisierung in einem CrowdRE-Setting unterstützen kann. Der Fokus lag dabei darauf, ob ChatGPT die Priorisierung von Anforderungen in die Kano-Kategorien erleichtert, den Ablauf beschleunigt und allgemein unterstützen kann. Um das herauszufinden, wurde eine Studie durchgeführt, bei der Stakeholder mit und ohne Unterstützung durch ChatGPT Anforderungen priorisierten und Begründungen formulierten, wobei diese Anforderungen mittels eines Visionvideos entstanden sind. Die Ergebnisse zeigen, dass der Einsatz von ChatGPT den Gesamtprozess unerwarteterweise eher verlängert hat. Gründe für diese Verlängerung könnten einerseits die zeitaufwendige Interaktion mit ChatGPT, das Lesen und Interpretieren der generierten Antworten sowie technische Faktoren, wie eine mögliche Verzögerung durch die Internetverbindung, gewesen sein. Ein weiterer wichtiger Aspekt der Untersuchung war die subjektive Wahrnehmung der Beteiligten hinsichtlich der Schwierigkeit des Priorisierungsprozesses. Hier zeigte sich, dass Stakeholder, die mit ChatGPT arbeiteten, den Prozess als etwas leichter empfanden als jene, die traditionelle Methoden nutzten. Sowohl in Bezug auf die benötigte Zeit als auch die empfundene Schwierigkeit konnte jedoch kein signifikanter Unterschied zwischen den beiden Gruppen festgestellt werden. Dennoch äußerten sich die Stakeholder positiv über die unterstützende Funktion des Chatbots. Zusammenfassend lässt sich vermuten, dass ChatGPT positiven Einfluss haben kann, jedoch weitere Studien klarere Ergebnisse liefern würden.

## 7.2 Ausblick

Zukünftige Studien könnten mit größeren Stichproben und einem längeren Untersuchungszeitraum weitere, signifikante Ergebnisse liefern, vor allem in Bezug auf das Potenzial von LLMs zur Entlastung von Stakeholdern. Insbesondere könnte der Einsatz neuer Versionen von ChatGPT oder anderer LLMs zu einer Verbesserung der Effizienz bei der Anforderungspriorisierung führen. Weitere Forschungsarbeiten könnten außerdem Priorisierungsergebnisse und Begründungen beider Gruppen analysieren, um die Qualität der von ChatGPT generierten Antworten genauer zu bewerten.

# Anhang A

# Studie

## A.1 Video


Abbildung A.1: Screenshot zum Visionvideo

## A.2 Onlineumfrage

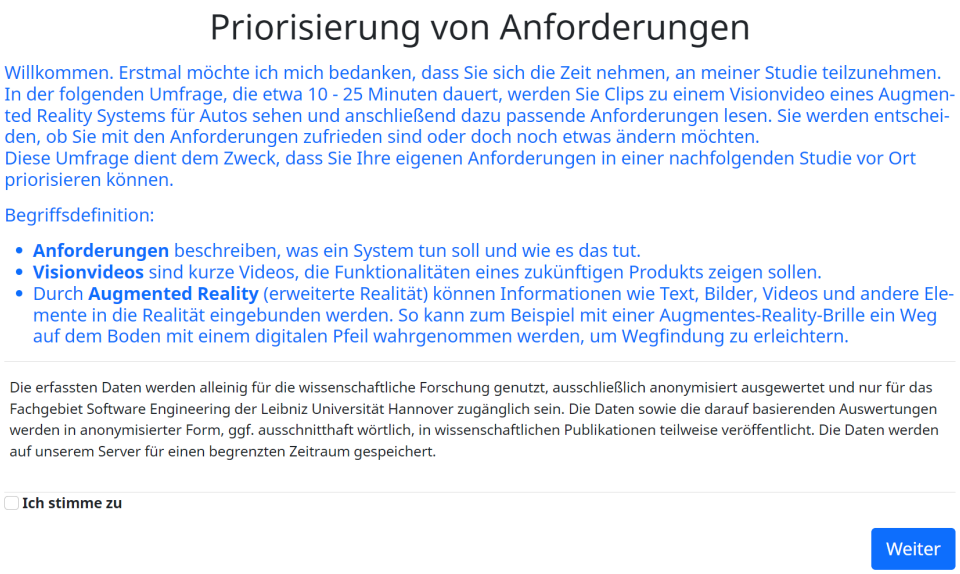
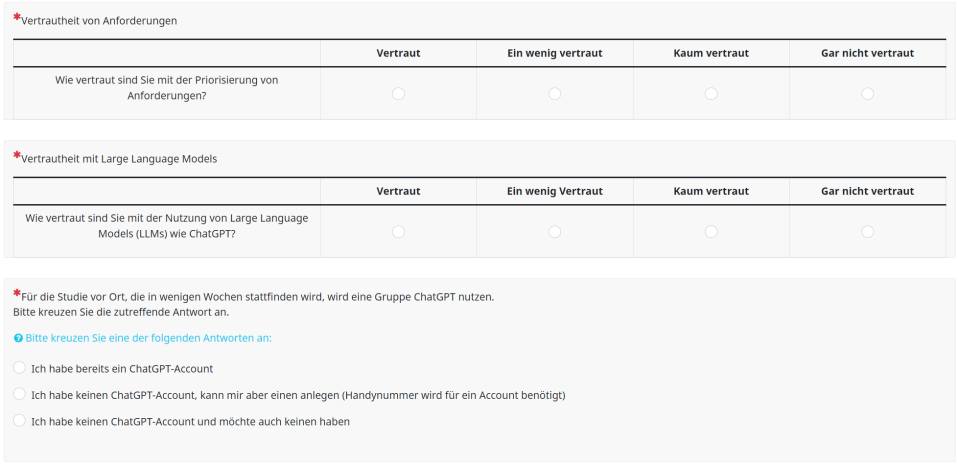



Abbildung A.2: Einwilligungserklärung der Onlineumfrage


Abbildung A.3: Demografische Daten 1



*Vertrautheit von Anforderungen

| | Vertraut | Ein wenig vertraut | Kaum vertraut | Gar nicht vertraut |
|---|---|---|---|---|
| Wie vertraut sind Sie mit der Priorisierung von Anforderungen? | | | | |

*Vertrautheit mit Large Language Models

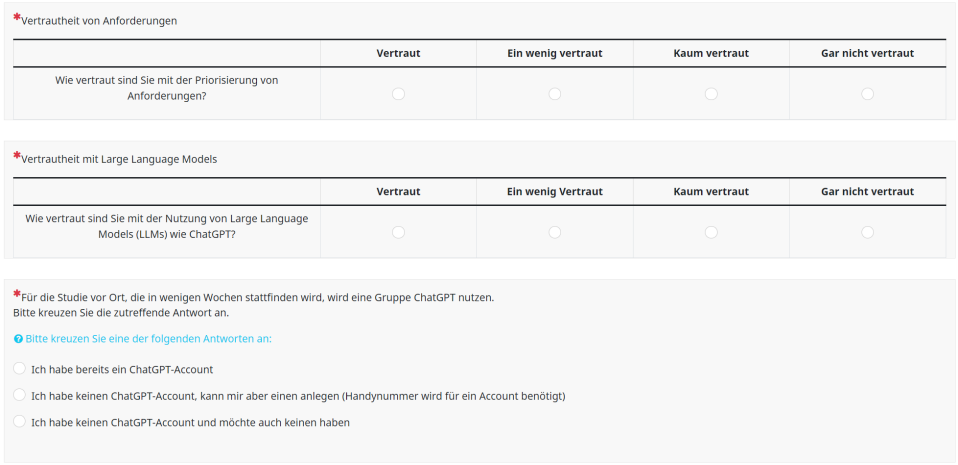
| | Vertraut | Ein wenig Vertraut | Kaum vertraut | Gar nicht vertraut |
|---|---|---|---|---|
| Wie vertraut sind Sie mit der Nutzung von Large Language Models (LLMs) wie ChatGPT? | | | | |

*Für die Studie vor Ort, die in wenigen Wochen stattfinden wird, wird eine Gruppe ChatGPT nutzen.
Bitte kreuzen Sie die zutreffende Antwort an.

Bitte kreuzen Sie eine der folgenden Antworten an:

- Ich habe bereits ein ChatGPT-Account
- Ich habe keinen ChatGPT-Account, kann mir aber einen anlegen (Handynummer wird für ein Account benötigt)
- Ich habe keinen ChatGPT-Account und möchte auch keinen haben

Abbildung A.4: Demografische Daten 2



*Vertrautheit von Anforderungen

| | Vertraut | Ein wenig vertraut | Kaum vertraut | Gar nicht vertraut |
|---|---|---|---|---|
| Wie vertraut sind Sie mit der Priorisierung von Anforderungen? | | | | |

*Vertrautheit mit Large Language Models

| | Vertraut | Ein wenig Vertraut | Kaum vertraut | Gar nicht vertraut |
|---|---|---|---|---|
| Wie vertraut sind Sie mit der Nutzung von Large Language Models (LLMs) wie ChatGPT? | | | | |

*Für die Studie vor Ort, die in wenigen Wochen stattfinden wird, wird eine Gruppe ChatGPT nutzen.
Bitte kreuzen Sie die zutreffende Antwort an.

Bitte kreuzen Sie eine der folgenden Antworten an:

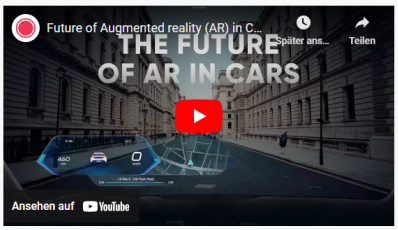
- Ich habe bereits ein ChatGPT-Account
- Ich habe keinen ChatGPT-Account, kann mir aber einen anlegen (Handynummer wird für ein Account benötigt)
- Ich habe keinen ChatGPT-Account und möchte auch keinen haben

Abbildung A.5: Demografische Daten 3

Visionvideo

Bitte schauen Sie sich zuerst das komplette Visionvideo "Future of Augmented reality (AR) in Cars" an.

*




Bitte kreuzen Sie eine der folgenden Antworten an:

Ich habe das komplette Video bis zum Ende geschaut

Abbildung A.6: Komplettes Visionvideo

*Schauen Sie sich zuerst den ersten Abschnitt an:


Jetzt folgen die Anforderungen zu dem Abschnitt. Bitte lesen Sie sich diese aufmerksam durch und entscheiden, ob Sie mit den Anforderungen zufrieden sind.

| | Ich stimme zu | Ich stimme nicht zu |
|---|---|---|
| Das Augmented-Reality-System im Auto muss fähig sein, die Daten des Navis zu empfangen (1.1) | | |
| Das Augmented-Reality-System des Autos muss fähig sein, auf der Windschutzscheibe eine Erinnerung, die Sonnenbrille anzuziehen, anzuzeigen (1.2) | | |
| Das Augmented-Reality-System des Autos muss fähig sein, auf der Windschutzscheibe eine Musikanzeige anzuzeigen (1.3) | | |
| Das Augmented-Reality-System des Autos muss dem Fahrer die Möglichkeit bieten, auf der Windschutzscheibe eine ausgewählte Strecke mit der Dauer der Fahrt anzuzeigen (1.4) | | |
| Das Augmented-Reality-System des Autos muss dem Fahrer die Möglichkeit bieten, auf der Windschutzscheibe die Position des Autos auf der Karte anzuzeigen (1.5) | | |
| Das Augmented-Reality-System des Autos muss dem Fahrer die Möglichkeit bieten, auf der Windschutzscheibe die Uhrzeit, Temperatur und Geschwindigkeit anzuzeigen (1.6) | | |
| Das Augmented-Reality-System des Autos muss komplette Funktionsfähigkeit innerhalb 5 Sekunden nach Start des Motors garantieren (1.7) | | |
| Das Augmented-Reality-System des Autos sollte in allen in Deutschland zugelassenen PKWs funktionieren (1.8) | | |

Falls Sie eigene Anforderungen zum Videoabschnitt haben, die hier nicht aufgelistet sind, können Sie diese im unteren Textfeld nummeriert aufschreiben.

Abbildung A.7: Erster Abschnitt des Visionvideos

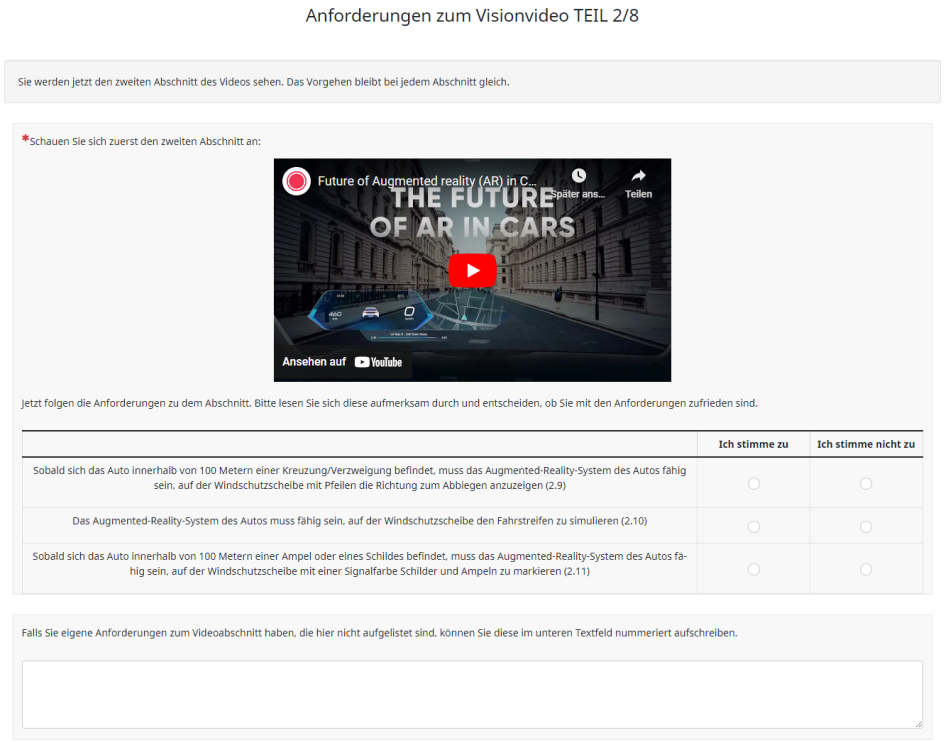



Abbildung A.8: Zweiter Abschnitt des Visionvideos

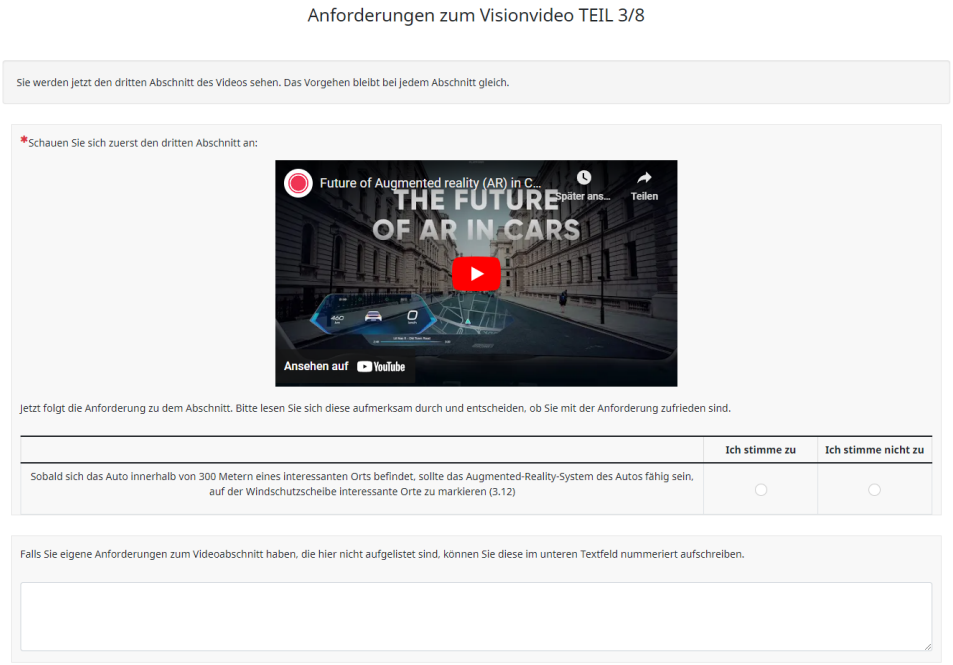
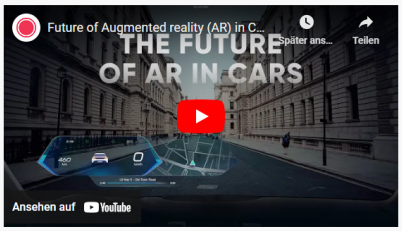



Abbildung A.9: Dritter Abschnitt des Visionvideos

Anforderungen zum Visionvideo TEIL 4/8

Sie werden jetzt den vierten Abschnitt des Videos sehen. Das Vorgehen bleibt bei jedem Abschnitt gleich.

*Schauen Sie sich zuerst den vierten Abschnitt an:




Jetzt folgen die Anforderungen zu dem Abschnitt. Bitte lesen Sie sich diese aufmerksam durch und entscheiden, ob Sie mit den Anforderungen zufrieden sind.

| | Ich stimme zu | Ich stimme nicht zu |
|---|---|---|
| Falls eine Ampel rot oder grün zeigt und sich das Auto innerhalb von 20 Metern einer Ampel befindet, muss das Augmented-Reality-System des Autos fähig sein, auf der Windschutzscheibe die verbleibende Zeit des Zustands der Ampel in Sekunden anzuzeigen (4.13) | ○ | ○ |
| Sobald sich das Auto innerhalb von 100 Metern einer Ampel befindet, muss das Augmented-Reality-System des Autos fähig sein, auf der Windschutzscheibe, mit der dementsprechenden Farbe, den Zustand der Ampel zu markieren (4.14) | ○ | ○ |
| Sobald sich das Auto innerhalb von 50 Metern eines Zebrastreifens befindet, muss das Augmented-Reality-System des Autos fähig sein, auf der Windschutzscheibe diesen Zebrastreifen zu markieren (4.15) | ○ | ○ |
| Die Markierung des Zebrastreifens muss erfolgen, wenn ein Passant in spätestens 3 Sekunden den Zebrastreifen betreten wird (4.16) | ○ | ○ |
| Das Augmented-Reality-System des Autos muss eine Reaktionszeit von 10ms garantieren (4.17) | ○ | ○ |

Falls Sie eigene Anforderungen zum Videoabschnitt haben, die hier nicht aufgelistet sind, können Sie diese im unteren Textfeld nummeriert aufschreiben.

Abbildung A.10: Vierter Abschnitt des Visionvideos

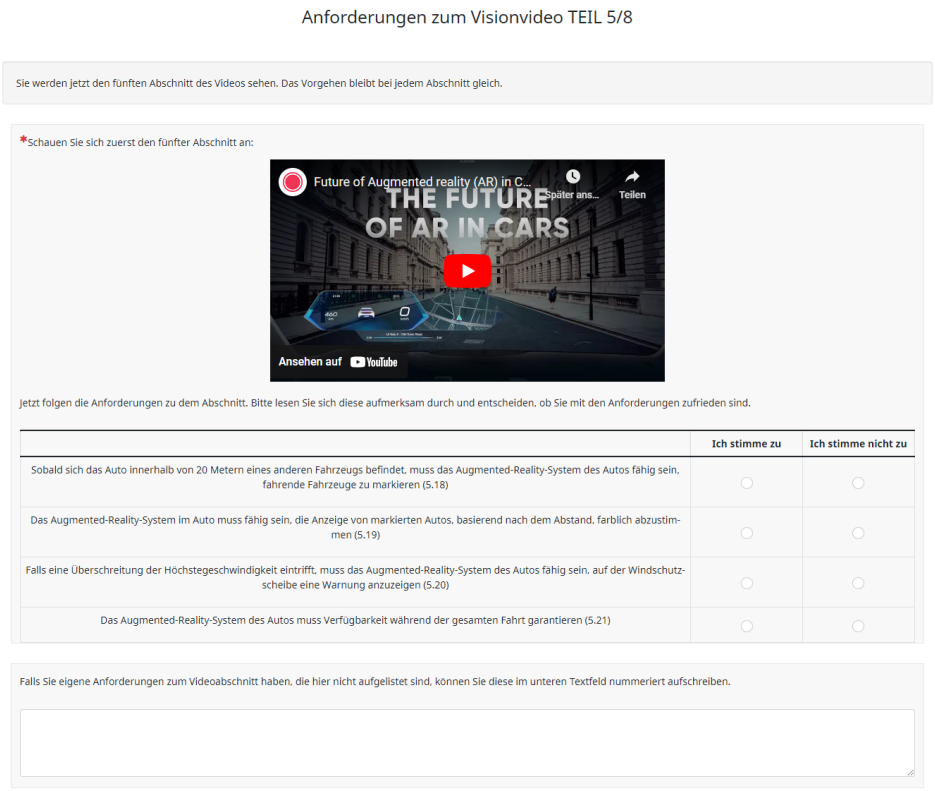



Abbildung A.11: Fünfter Abschnitt des Visionvideos

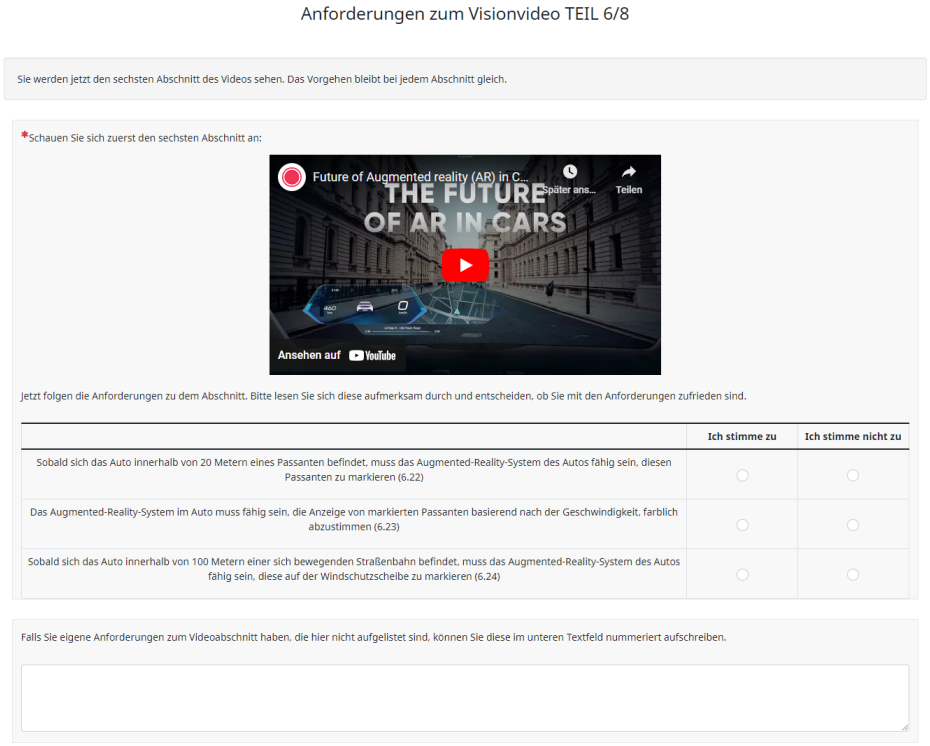



Abbildung A.12: Sechster Abschnitt des Visionvideos

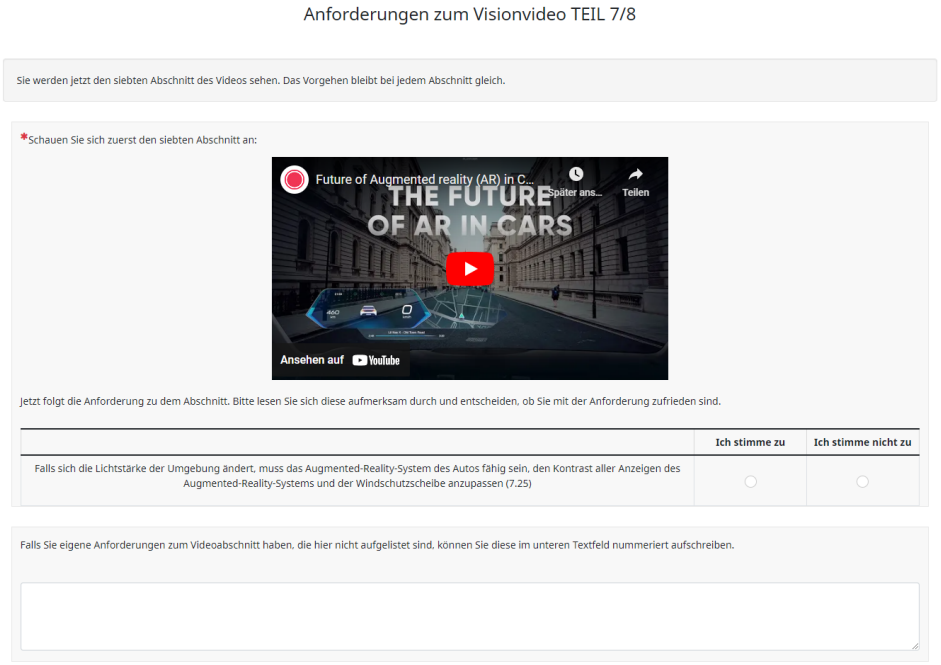



Abbildung A.13: Siebter Abschnitt des Visionvideos

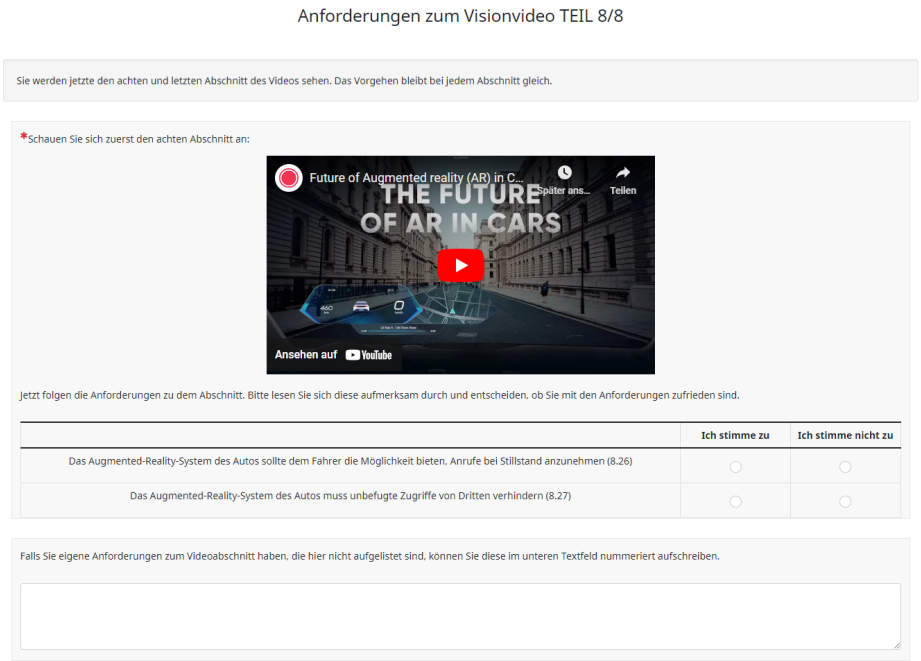



Abbildung A.14: Achter Abschnitt des Visionvideos

## A.3 Handouts

## Kano

Das Kano-Modell beschreibt den Zusammenhang zwischen Produkteigenschaften (Anforderungen) und der erwarteten Zufriedenheit von Kunden. Das Modell erlaubt es, Kundenwünsche zu erfassen und bei der Entwicklung zu berücksichtigen.

Es existieren 3 Ebenen von Merkmalen:

- Basis-Merkmale (muss):
  Diese Merkmale sind grundlegend. Kunden nehmen diese Merkmale als so selbstverständlich war, dass sie sich diesen erst bewusstwerden, wenn das Produkt diese Merkmale nicht erfüllt.
  Bei Erfüllung entsteht keine Zufriedenheit (da selbstverständlich), bei Nichterfüllung jedoch werden Nutzer unzufrieden.
  - Beispiel (Auto):
    - funktionierende Airbags
    - Rostschutz
- Leistungsmerkmale (soll):
  Sind dem Kunden bewusst (werden also explizit verlangt). Sie beseitigen Unzufriedenheit oder schaffen Zufriedenheit abhängig von Ausmaß der Erfüllung. Vorhandensein schafft Zufriedenheit, Fehlen schafft Unzufriedenheit.
  - Beispiel (Auto):
    - lange Lebensdauer
    - hohe Beschleunigung
- Begeisterungsmerkmale (kann):
  Merkmale, die Kunden nicht unbedingt erwarten. Sie heben das Produkt von der Konkurrenz ab und begeistern Kunden.
  Vorhandensein schafft Zufriedenheit, Fehlen schafft keine Unzufriedenheit.
  - Beispiel (Auto)
    - Sonderausstattung
    - besonderes Design

Erklärung und Beispiele entnommen aus https://six-sigma.com/lexikon/kano-analyse/ und https://de.wikipedia.org/wiki/Kano-Modell

Abbildung A.15: Kano-Handout

## ChatGPT Handout

Funktionen

- In die TextBox können Sie Ihre Nachricht an den Chatbot **schreiben**.


- Mit dem Pfeil rechts neben der TextBox können Sie Ihre **Nachricht abschicken**. Alternativ können Sie mit der Return-Taste auf der Tastatur dasselbe erzielen.
- Wenn Sie mit der Maus über Ihre eigene Nachricht gehen (ohne zu klicken), erscheint ein Stift-Symbol. Mit dem können Sie ihre **Nachricht bearbeiten**.
- Bei einer Antwort von ChatGPT können Sie mit den 2 Pfeilen, die unterhalb der Nachricht ist, einen Kreis bilden, ChatGPT's **Antwort neu generieren**.
- Bei einer Antwort von ChatGPT können Sie mit den 2 übereinanderliegenden Vierecken die **Nachricht kopieren**.
- Bei einer Antwort von ChatGPT können Sie außerdem, sofern Sie die Antwort neu generiert oder Ihre eigene Nachricht verändert haben, zwischen allen bereits generierten **Antworten wechseln**. Nutzen Sie dafür die 2 Pfeile 

, die ganz links unterhalb der Antwort angezeigt werden.

Rollen einnehmen

- Um präzisere Antworten zu erhalten, können Sie ChatGPT eine bestimmte Rolle zuweisen. In dieser Studie wäre es sinnvoll, dass ChatGPT die Rolle eines Anforderungsingenieurs erhält.
  Beispiel: "Nimm von jetzt an die Rolle eines Anforderungsingenieurs ein."
  Danach können Sie das Szenario beschreiben und Ihre Anforderungen priorisieren.

Abbildung A.16: ChatGPT-Handout Seite 1

Weitere Tipps

- Sie können ChatGPT für alle Verständnisfragen verwenden. Es kann Ihnen erklären, was genau das Kano-Modell ist, welche Kategorien es gibt und Tipps geben, wie nach Kano priorisiert werden sollte.
- Zusätzlich können Sie ChatGPT nach Definitionen von Begriffen fragen.
  Beispiel: “Definiere Augmented Reality und gib kurze Anwendungsbeispiele an”.
- Sie können kleine Tippfehler ignorieren, Höflichkeiten weglassen und mit ChatGPT so chatten, wie mit einem Menschen.
- Wenn Sie eine Antwort von ChatGPT anzweifeln, können Sie ihn dazu bringen, seine Antwort nochmal zu überdenken.
  Z.B.:
  “Mit dem Hintergrund der Begründung, die du mir gerade gegeben hast, überdenke nochmal deine Priorisierung”
  oder
  “Bist du dir sicher, dass die gerade behandelte Anforderung wirklich ein Basismerkmal ist? Wenn ja, Begründe. Sonst überdenke deine Antwort.”
- Sie können sich **jederzeit gegen ChatGPT’s Antworten entscheiden**. Falls Sie Ihre Priorisierung und Ihre Begründung für die Priorisierung besser finden, können Sie diese gerne übernehmen.

Beispiel

- “Nimm ab jetzt die Rolle eines Anforderungsingenieurs an. Du sollst folgende Anforderung nach dem Kano-Modell priorisieren:
  '[Hier Anforderung reinkopieren]’.
  Begründe anschließend deine Priorisierung”**.**
  (Sie müssen die Beispiele im Handout nicht übernehmen)

Abbildung A.17: ChatGPT-Handout Seite 2

*Juli 2024*

## Überblick über die Nutzerstudie: „Priorisierung von Anforderungen mit und ohne ChatGPT“

**Fachgebiet Software Engineering, Leibniz Universität Hannover**

Bitte lesen Sie diesen Überblick sorgfältig durch. Dieser Überblick dient dazu, Ihnen die Studie vorzustellen und Sie auf Ihre Rechte als freiwillige*r Studienteilnehmer*in hinzuweisen. Bitte fragen Sie nach, wenn etwas unklar sein sollte.

Wir erheben Daten von mehreren Teilnehmenden. Die in der Studie gesammelten Daten helfen uns, den Vergleich von ChatGPT zu traditionellen Methoden im Kontext von Anforderungspriorisierung zu evaluieren und besser zu verstehen.

**Sie müssen folgende Voraussetzungen für die Teilnahme erfüllen:**

1. **Sie haben an der ersten Umfrage “Priorisierung von Anforderungen” teilgenommen.**

### Ablauf und Dauer

Ihre Aufgabe ist die Priorisierung von Anforderungen nach dem Kano-Modell. Hierfür können Sie das Formular “Umfrage.pdf” nutzen.
Weitere Infos zum Kano-Modell und Priorisierung finden Sie im Handout.

Gehen Sie diesen Ablauf für jede Anforderung durch:
Betrachten Sie eine Anforderung und ordnen Sie diese in eine der 3 Kano-Kategorien. Anschließend begründen Sie, wieso die Anforderung genauso priorisiert worden ist.

Gibt es keine Anforderungen mehr, können Sie den restlichen Teil der Umfrage.pdf ausfüllen.
Für die Studie dürfen Sie folgenden nutzen: PDF mit Ihren Anforderungen, Editor, Handouts, Umfrage.pdf.
Die verwendete Sprache in der Studie ist **Deutsch**. Die Studie wird insgesamt **ca. 45 Minuten** dauern.

### Datenschutz und Datenspeicherung

Die von Ihnen zur Verfügung gestellten Daten werden ausschließlich anonym und ohne Rückschlüsse auf einzelne Personen ausgewertet. Vor der Verarbeitung der Daten erfolgt eine umfangreiche Anonymisierung. Die Daten werden auf unserem Server für einen begrenzten Zeitraum gespeichert. Die Daten sowie die darauf basierenden Auswertungen werden im Rahmen wissenschaftlicher Publikationen in anonymisierter Form teilweise veröffentlicht.

**Bitte lesen Sie sich nun sorgfältig die Einverständniserklärung auf der nächsten Seite und die dort beschriebenen Hinweise durch.**

Abbildung A.18: Einverständniserklärung und Ablauf für die traditionelle Gruppe Seite 1

*Juli 2024*

**Einverständniserklärung zu der Nutzerstudie:**
**„Priorisierung von Anforderungen mit und ohne ChatGPT“**
**Fachgebiet Software Engineering, Leibniz Universität Hannover**

Diese Studie wird von dem Fachgebiet Software Engineering der Leibniz Universität Hannover durchgeführt. Die Studie wird von Yevgen Tytov, einem Studenten für seine Bachelorarbeit „Evaluation von LLMs in der Anforderungspriorisierung ausgehend von Feedback zu Visionsvideos“ am Fachgebiet Software Engineering, durchgeführt.
Ich habe den Überblick über die Studie gelesen und verstanden. Ich nehme freiwillig und ohne Vergütung an dieser Studie teil. Ich habe das Recht, die Teilnahme jederzeit und ohne Angabe von Gründen abzubrechen.

Diese Einwilligung ist freiwillig. Ich kann sie ohne Angabe von Gründen verweigern, ohne dass ich deswegen Nachteile zu befürchten hätte.

**1. Datenerfassung, -speicherung und -verwendung**

Ich wurde darüber informiert, dass alles Geschriebene erfasst, elektronisch für einen begrenzten Zeitraum gespeichert und zur Auswertung der Nutzerstudie herangezogen wird. Die erfassten Daten werden alleinig für die wissenschaftliche Forschung genutzt, ausschließlich anonymisiert ausgewertet und nur für das Fachgebiet Software Engineering der Leibniz Universität Hannover zugänglich sein. Die Daten sowie die darauf basierenden Auswertungen werden in anonymisierter Form, ggf. ausschnitthaft wörtlich, in wissenschaftlichen Publikationen teilweise veröffentlicht.

Sie haben gemäß Datenschutz gegenüber dem Informationsträger das Recht auf Auskunft, Berichtigung sowie Löschung Ihrer personenbezogenen Daten. Sie können diese Einwilligungserklärung jederzeit schriftlich widerrufen. Nach erfolgtem Widerruf werden Ihre personenbezogenen Daten gelöscht und ab diesem Zeitpunkt für keine weiteren Publikationen mehr verwendet.

**2. Unterschrift**

Ich habe den Überblick über die Studie gelesen und verstanden. **Ich erfülle die genannten Voraussetzungen für die Teilnahme.** Ich bin mit den aufgeführten Punkten der Einverständniserklärung einverstanden. Ich nehme freiwillig und ohne Vergütung an dieser Studie teil.

_____
Nachname, Vorname (in Blockschrift)

_____
Ort, Datum, Unterschrift

Abbildung A.19: Einverständniserklärung und Ablauf für die traditionelle Gruppe Seite 2

*Juli 2024*

## Überblick über die Nutzerstudie: „Priorisierung von Anforderungen mit und ohne ChatGPT“

**Fachgebiet Software Engineering, Leibniz Universität Hannover**

Bitte lesen Sie diesen Überblick sorgfältig durch. Dieser Überblick dient dazu, Ihnen die Studie vorzustellen und Sie auf Ihre Rechte als freiwillige*r Studienteilnehmer*in hinzuweisen. Bitte fragen Sie nach, wenn etwas unklar sein sollte.

Wir erheben Daten von mehreren Teilnehmenden. Die in der Studie gesammelten Daten helfen uns, den Vergleich von ChatGPT zu traditionellen Methoden im Kontext von Anforderungspriorisierung zu evaluieren und besser zu verstehen.

**Sie müssen folgende Voraussetzungen für die Teilnahme erfüllen:**

1. **Sie haben an der ersten Umfrage “Priorisierung von Anforderungen” teilgenommen.**

### Ablauf und Dauer

Ihre Aufgabe ist die Priorisierung dieser Anforderungen nach dem Kano-Modell. Hierfür steht ihnen das Large Language Model “ChatGPT” zur Verfügung. Sie können ChatGPT nutzen, um Hilfe bei Verständnisfragen und der Priorisierung zu bekommen. Natürlich müssen Sie sich nicht auf ChatGPT verlassen und sollen bei der Priorisierung Ihre eigene Meinung einbringen.
Weitere Infos zum Kano-Modell, ChatGPT und Priorisierung finden Sie im Handout.

Wiederholen Sie folgende Schritte, bis alle Anforderungen abgearbeitet sind.
**Schritt 1**: Betrachten Sie eine Anforderung und ordnen Sie diese in eine der 3 Kano-Kategorien. Sie dürfen dafür ChatGPT verwenden. Anschließend begründen Sie, wieso die Anforderung genauso priorisiert worden ist. Sie können die Begründung selbst aufstellen oder von ChatGPT ganz oder teilweise kopieren. Sind Sie mit dem Ergebnis zufrieden, können Sie mit Schritt 2 fortfahren. Sind Sie nicht zufrieden, haben Sie die Gelegenheit, die Regenerate-Funktion von ChatGPT zu nutzen oder ChatGPT zu sagen, dass er seine Antwort überdenken soll. Wenn Sie der Meinung sind, dass ChatGPT keine annehmbare Lösung gefunden hat, können Sie die Anforderung ohne Hilfe priorisieren.
**Schritt 2**: Wechseln Sie jetzt zur Umfrage.pdf. Dort geben Sie für die Anforderung die Priorisierung an, ob es von ChatGPT übernommen wurde und schreiben dort die Begründung rein.

Gibt es keine Anforderungen mehr, können Sie den restlichen Teil der Umfrage.pdf ausfüllen. Hierfür brauchen Sie ChatGPT nicht.
Für die Studie dürfen Sie folgenden nutzen: PDF mit Ihren Anforderungen, ChatGPT, Editor, Handouts, Umfrage.pdf.
Die verwendete Sprache in der Studie ist **Deutsch**. Die Studie wird insgesamt **ca. 45 Minuten** dauern.

### Datenschutz und Datenspeicherung

Die von Ihnen zur Verfügung gestellten Daten werden ausschließlich anonym und ohne Rückschlüsse auf einzelne Personen ausgewertet. Vor der Verarbeitung der Daten erfolgt eine umfangreiche Anonymisierung. Die Daten werden auf unserem Server für einen begrenzten Zeitraum gespeichert. Die Daten sowie die darauf basierenden Auswertungen werden im Rahmen wissenschaftlicher Publikationen in anonymisierter Form teilweise veröffentlicht.
**Bitte lesen Sie sich nun sorgfältig die Einverständniserklärung auf der nächsten Seite und die dort beschriebenen Hinweise durch.**

Abbildung A.20: Einverständniserklärung und Ablauf für die ChatGPT-Gruppe Seite 1

*Juli 2024*

**Einverständniserklärung zu der Nutzerstudie:**
**„Priorisierung von Anforderungen mit und ohne ChatGPT“**
**Fachgebiet Software Engineering, Leibniz Universität Hannover**

Diese Studie wird von dem Fachgebiet Software Engineering der Leibniz Universität Hannover durchgeführt. Die Studie wird von Yevgen Tytov, einem Studenten für seine Bachelorarbeit „Evaluation von LLMs in der Anforderungspriorisierung ausgehend von Feedback zu Visionsvideos“ am Fachgebiet Software Engineering, durchgeführt.
Ich habe den Überblick über die Studie gelesen und verstanden. Ich nehme freiwillig und ohne Vergütung an dieser Studie teil. Ich habe das Recht, die Teilnahme jederzeit und ohne Angabe von Gründen abzubrechen.

Diese Einwilligung ist freiwillig. Ich kann sie ohne Angabe von Gründen verweigern, ohne dass ich deswegen Nachteile zu befürchten hätte.

**1. Datenerfassung, -speicherung und -verwendung**

Ich wurde darüber informiert, dass alles Geschriebene erfasst, elektronisch für einen begrenzten Zeitraum gespeichert und zur Auswertung der Nutzerstudie herangezogen wird. Die erfassten Daten werden alleinig für die wissenschaftliche Forschung genutzt, ausschließlich anonymisiert ausgewertet und nur für das Fachgebiet Software Engineering der Leibniz Universität Hannover zugänglich sein. Die Daten sowie die darauf basierenden Auswertungen werden in anonymisierter Form, ggf. ausschnitthaft wörtlich, in wissenschaftlichen Publikationen teilweise veröffentlicht.

Sie haben gemäß Datenschutz gegenüber dem Informationsträger das Recht auf Auskunft, Berichtigung sowie Löschung Ihrer personenbezogenen Daten. Sie können diese Einwilligungserklärung jederzeit schriftlich widerrufen. Nach erfolgtem Widerruf werden Ihre personenbezogenen Daten gelöscht und ab diesem Zeitpunkt für keine weiteren Publikationen mehr verwendet.

**2. Unterschrift**

Ich habe den Überblick über die Studie gelesen und verstanden. **Ich erfülle die genannten Voraussetzungen für die Teilnahme.** Ich bin mit den aufgeführten Punkten der Einverständniserklärung einverstanden. Ich nehme freiwillig und ohne Vergütung an dieser Studie teil.

_____
Nachname, Vorname (in Blockschrift)

_____
Ort, Datum, Unterschrift

Abbildung A.21: Einverständniserklärung und Ablauf für die ChatGPT-Gruppe Seite 2

## A.4 Fomular

Es folgen Fragen zu Ihren Erfahrungen, die Sie in der Studie gemacht haben.
Bitte beantworten Sie die Fragen ohne Hilfe von ChatGPT.

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  die Entscheidungen von ChatGPT, wie es Anforderungen nach Kano priorisiert hat, waren nachvollziehbar.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  das Kano-Modell war für mich verständlich.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  ChatGPT konnte mir helfen, das Kano-Modell besser zu verstehen.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  es gab Momente, wo ich Schwierigkeiten hatte, eine Anforderung zu priorisieren.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  ChatGPT war mir bei der Priorisierung von Anforderungen nach Kano eine Hilfe.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

Abbildung A.22: Frageteil des Formulars der ChatGPT-Gruppe Seite 1

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  es gab Momente, wo ich Schwierigkeiten hatten, eine Begründung für eine Priorisierung zu finden.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  ChatGPT konnte mir bei der Begründung der Priorisierungen helfen.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

Abbildung A.23: Frageteil des Formulars der ChatGPT-Gruppe Seite 2

Es folgen Fragen zu Ihren Erfahrungen, die Sie in der Studie gemacht haben.

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  das Kano-Modell war für mich verständlich.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  es gab Momente, wo ich Schwierigkeiten hatte, eine Anforderung zu priorisieren.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

- Wie sehr stimmen Sie mit der folgenden Aussage überein:
  es gab Momente, wo ich Schwierigkeiten hatten, eine Begründung für eine Priorisierung zu finden.

| Voll und ganz | Ein wenig | neutral | kaum | Gar nicht |
|---|---|---|---|---|
| ○ | ○ | ○ | ○ | ○ |

Abbildung A.24: Frageteil des Forumlars der traditionellen Gruppe

# Literaturverzeichnis


[1] Openai. `https://platform.openai.com/docs/models/gpt-4o`. Accessed: 2024-08-18.

[2] P. Achimugu, A. Selamat, R. Ibrahim, and M. N. Mahrin. A systematic literature review of software requirements prioritization research. *Information and software technology*, 56(6):568–585, 2014.

[3] M. Arifianto and I. Izzudin. Students' acceptance of discord as an alternative online learning media. *International Journal of Emerging Technologies in Learning (iJET)*, 16(20):179–195, 2021.

[4] M. I. Babar, M. Ramzan, and S. A. Ghayyur. Challenges and future trends in software requirements prioritization. In *International conference on computer networks and information technology*, pages 319–324. IEEE, 2011.

[5] V. R. Basili and H. D. Rombach. The tame project: Towards improvement-oriented software environments. *IEEE Transactions on software engineering*, 14(6):758–773, 1988.

[6] M. Binder, A. Vogt, A. Bajraktari, and A. Vogelsang. Automatically classifying kano model factors in app reviews. In *International Working Conference on Requirements Engineering: Foundation for Software Quality*, pages 245–261. Springer, 2023.

[7] M. Broy, E. Geisberger, J. Kazmeier, A. Rudorfer, and K. Beetz. Ein requirements-engineering-referenzmodell. *Informatik-Spektrum*, 30:127–142, 2007.

[8] B. Chen, K. Chen, S. Hassani, Y. Yang, D. Amyot, L. Lessard, G. Mussbacher, M. Sabetzadeh, and D. Varró. On the use of gpt-4 for creating goal models: an exploratory study. In *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, pages 262–271. IEEE, 2023.

[9] A. Dryaev. Benutzung von large-language-models für das formulieren der anforderungen aus visionsvideos, December 2023.

Available at https://www.pi.uni-hannover.de/fileadmin/pi/se/Stud-Arbeiten/2023/BA_Dryaev2023.pdf#page=72&zoom=100,342,944.

[10] C. Duan, P. Laurent, J. Cleland-Huang, and C. Kwiatkowski. Towards automated requirements prioritization and triage. *Requirements engineering*, 14:73–89, 2009.

[11] I. Etikan, S. A. Musa, R. S. Alkassim, et al. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1):1–4, 2016.

[12] A. Fantechi, S. Gnesi, L. Passaro, and L. Semini. Inconsistency detection in natural language requirements using chatgpt: a preliminary evaluation. In *2023 IEEE 31st International Requirements Engineering Conference (RE)*, pages 335–340. IEEE, 2023.

[13] M. Glinz. On non-functional requirements. In *15th IEEE international requirements engineering conference (RE 2007)*, pages 21–26. IEEE, 2007.

[14] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini, et al. The crowd in requirements engineering: The landscape and challenges. *IEEE software*, 34(2):44–52, 2017.

[15] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea Preprints*, 2023.

[16] O. Karras, K. Schneider, and S. A. Fricker. Representing software project vision by means of video: a quality model for vision videos. *journal of Systems and Software*, 162:110479, 2020.

[17] M. Krishna, B. Gaur, A. Verma, and P. Jalote. Using llms in software requirements specifications: An empirical evaluation. *arXiv preprint arXiv:2404.17842*, 2024.

[18] C. Kultanan, H.-A. Crostack, and R. Refflinghaus. Implementation of kano methodology through various stakeholder requirements. In *Proceedings of the 7th Asia Pacific Industrial Engineering and Management Systems Conference*, pages 17–20, 2006.

[19] P. Madzík. Increasing accuracy of the kano model–a case study. *Total Quality Management & Business Excellence*, 29(3-4):387–409, 2018.

[20] D. Mairiza, D. Zowghi, and N. Nurmuliani. An investigation into the notion of non-functional requirements. In *Proceedings of the 2010 ACM symposium on applied computing*, pages 311–317, 2010.

[21] G. Marvin, N. Hellen, D. Jjingo, and J. Nakatumba-Nabende. Prompt engineering in large language models. In *International conference on data intelligence and cognitive informatics*, pages 387–402. Springer, 2023.

[22] R. A. F. PIAM. Erhebung und validierung von testbaren anforderungen durch visionvideos, June 2023. Available at `https://www.pi.uni-hannover.de/fileadmin/pi/se/Stud-Arbeiten/2023/MA_FOKAM_PIAM.pdf`.

[23] O. Pieczul, S. Foley, and M. E. Zurko. Developer-centered security and the symmetry of ignorance. In *Proceedings of the 2017 New Security Paradigms Workshop*, pages 46–56, 2017.

[24] E. Reinking and M. Becker. Einsatzmöglichkeiten von ki in unternehmen-zeitnah erreichbare personalisierung von large language models (wie chatgpt) in zeiten der industrie 5.0. Technical report, IUCF Working Paper, 2023.

[25] K. Ruan, X. Chen, and Z. Jin. Requirements modeling aided by chatgpt: An experience in embedded systems. In *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, pages 170–177. IEEE, 2023.

[26] C. Rupp et al. *Requirements-Engineering und-Management: Das Handbuch für Anforderungen in jeder Situation*. Carl Hanser Verlag GmbH Co KG, 2020.

[27] M. A. Sami, Z. Rasheed, M. Waseem, Z. Zhang, T. Herda, and P. Abrahamsson. Prioritizing software requirements using large language models. *arXiv preprint arXiv:2405.01564*, 2024.

[28] A. Shahin and M. Zairi. Kano model: A dynamic approach for classifying and prioritising requirements of airline travellers with three case studies on international airlines. *Total Quality Management*, 20(9):1003–1028, 2009.

[29] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, A. Wesslén, et al. *Experimentation in software engineering*, volume 236. Springer, 2012.

[30] D. Xie, B. Yoo, N. Jiang, M. Kim, L. Tan, X. Zhang, and J. S. Lee. Impact of large language models on generating software specifications. *arXiv preprint arXiv:2306.03324*, 2023.

[31] Y. S. Yang. Anforderungspriorisierung ausgehend von emotionen in visionsvideos, May 2023. Available at `https://www.pi.uni-hannover.de/fileadmin/pi/se/Stud-Arbeiten/2023/BA_Yang.pdf`.

[32] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, page 100211, 2024.

[33] X. Zhang et al. Personagen: a tool for generating personas from user feedback (2023).

[34] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.