

KI generierte Zusammenfassungen von Forumsdiskussionen über Marktwerte: Eine Bewertung von zentralen Aspekten

BACHELOR THESIS

Juni 2025

Autor
Fluri Elias

Betreuungsperson
Fred van den Anker

Praxispartner*in
Transfermarkt GmbH & Co. KG
Miriam Schweikert

Abstract:

Die Website Transfermarkt.de zeigt aktuelle Einschätzungen der Marktwerte von Fussballprofis. Diese Einschätzung basiert zu grossen Teilen auf öffentlichen Forumdiskussionen. Dabei tauscht sich eine aktive Community über die Leistung der Spieler, Potenzial, Position sowie weiteren Faktoren aus. Die Forumsbeiträge sind teilweise lang und erschweren die Nachvollziehbarkeit der wesentlichen Argumente, die zur Festlegung des Marktwerts geführt haben. Die vorliegende Bachelorarbeit beschäftigt sich mit der Frage, wie aktive Nutzerinnen und Nutzer von Transfermarkt.de eine prototypische durch Künstliche Intelligenz generierte Zusammenfassung einer Forumdiskussion über Marktwerte bewerten und welcher Änderungsbedarf besteht.

Für die qualitative Datenerhebung wurden halbstandardisierte Interviews mit zwölf aktiven Nutzerinnen und Nutzern geführt, transkribiert und mit MAXQDA 2024 inhaltsanalytisch ausgewertet. Die Bachelorarbeit ist theoriegeleitet aufgebaut. In der ausführlichen Literaturrecherche wurden relevante Aspekte für die Bewertung von KI-generierten Zusammenfassungen festgestellt. Die Aspekte Genauigkeit, Umfang, Vertrauen, Transparenz und Kontrolle, Stil und Format sowie Effizienz und Effektivität bilden die Basis für die Interviewfragen.

Diese Bachelorarbeit zeigt bedeutende Erkenntnisse für den Fortschritt von KI-generierten Zusammenfassungen auf der Website von Transfermarkt.de. Die Wichtigkeit der frühzeitigen Integration von Nutzerinnen und Nutzern in den Gestaltungsprozess wird hervorgehoben. Dies bietet einen praxisorientierten und zugleich theoretisch fundierteren Vorschlag für die konkrete Implementierung von KI-generierten Zusammenfassungen umfangreicher Forumdiskussionen.

Anzahl Zeichen im Bericht (ohne Anhang):

72'653

Inhaltsverzeichnis

1	Einleitung	1
1.1	Was ist Transfermarkt?.....	1
1.2	Problemstellung.....	1
1.3	Zielsetzung	2
1.4	Fragestellungen.....	2
1.5	Aufbau der Arbeit.....	2
1.6	Abgrenzung.....	3
2	Theorie.....	4
2.1	Human-Centered-Design-Prozess.....	4
2.2	Künstliche Intelligenz und Zusammenfassungen	5
2.3	Relevante Aspekte.....	5
2.3.1	Genauigkeit	6
2.3.2	Umfang.....	6
2.3.3	Vertrauen.....	6
2.3.4	Transparenz und Kontrolle.....	7
2.3.5	Stil und Format	7
2.3.6	Effizienz und Effektivität.....	7
3	Methoden	9
3.1	Untersuchungsplan.....	9
3.2	Halbstandardisierte Interviews.....	9
3.2.1	Erstellung des Leitfadens	10
3.2.2	Auswahl der Forumsbeiträge	11
3.2.3	Erstellung der prototypischen Zusammenfassung	11
3.2.4	Pre-Test Interviewleitfaden.....	11
3.2.5	Beschreibung der Stichprobe.....	12
3.2.6	Datenerhebung.....	12
3.3	Datenauswertung	12
3.3.1	Transkription.....	12
3.3.2	Kategoriensystem.....	13

3.3.3	Auswertung	15
4	Ergebnisse	16
4.1	Genauigkeit	16
4.1.1	Bewertung	16
4.1.2	Änderungsbedarf	18
4.2	Umfang.....	18
4.2.1	Bewertung	18
4.2.2	Änderungsbedarf	19
4.3	Vertrauen.....	20
4.3.1	Bewertung	20
4.3.2	Änderungsbedarf	22
4.4	Transparenz und Kontrolle	23
4.4.1	Bewertung	23
4.4.2	Änderungsbedarf	24
4.5	Stil und Format	25
4.5.1	Bewertung	25
4.5.2	Änderungsbedarf	26
4.6	Effizienz und Effektivität.....	27
4.6.1	Bewertung	27
4.7	Weitere Anwendungen	29
5	Diskussion.....	30
5.1	Zusammenfassung und Interpretation der Forschungsergebnisse.....	30
5.2	Kritische Würdigung und Fazit	31
5.3	Ausblick.....	32
6	Literaturverzeichnis	33
7	Abbildungsverzeichnis.....	36
8	Tabellenverzeichnis	36
9	Hilfsmittelverzeichnis.....	36
10	Anhangsverzeichnis	37

1 Einleitung

1.1 Was ist Transfermarkt?

Transfermarkt GmbH & Co. KG betreibt das Onlineportal Transfermarkt in verschiedenen Sprachen und Domains. Die Plattformen stellt eine breit gefächerte Auswahl von Informationen und Statistiken rund um den Fussball zur Verfügung. Dazu gehören Nachrichten zu Transfers im internationalen Fussball, Statistiken zu Leistungsdaten, Vereinshistorien oder Marktwerte. Dieser letztgenannte Marktwert hat innerhalb von Transfermarkt ein besonderer Stellenwert. Der Marktwert beschreibt den Wert, zu dem ein Spieler gehandelt wird, beziehungsweise den Wert, der einem Spieler zugeschrieben wird. Die Marktwerte werden durch eine freiwillige Community diskutiert und diese Diskussionen haben einen starken Einfluss auf die definitive Bewertung eines Spielers („Marktwert-Definition - So funktioniert´s - Marktwertanalyse - Forum | Seite 1“, 2021). Die Forummitglieder argumentieren und diskutieren fortlaufend in spezifischen online Foren auf Transfermarkt über den Marktwert eines Spielers. Sie beziehen sich bei ihrer Argumentation beispielsweise auf verschiedene Faktoren wie das Alter, die sportlichen Leistungen oder das Entwicklungspotenzial. Die Marktwerte werden in der Regel halbjährlich durch sogenannte Paten aktualisiert. Dabei handelt es sich um freiwillige Forummitglieder, die durch eine vorhergehendes Bewerbungsverfahren bestimmt wurden („Datenpflege-“How-to-Guide““, n. d.).

Die Transfermarkt-Marktwerte geniessen mittlerweile einen hohen medialen und Branchen spezifischen Stellenwert („20 Jahre Marktwerte: Vom ‚reinen Hobby-Projekt‘ zur relevanten Größe im Weltfußball“, 2024). So werden die Marktwerte regelmässig von bekannten Medien benutzt und als Teil der Berichterstattung genannt (deutschlandfunk.de, 2020). Die laufenden Diskussionen zu den Marktwerten sind öffentlich und somit für jede Person einsehbar.

1.2 Problemstellung

Die grosse Anzahl von Spielern und Marktwerten in der Datenbank führt zu einer unübersichtlichen Situation, wenn es um die Nachvollziehbarkeit der Entstehung der einzelnen Werte geht. Die Marktwerte werden derzeit mit einer Grafik, welche die historische Entwicklung aufzeigt, dargestellt. Um den aktuellen Marktwert eines Spielers nachvollziehen zu können, kann auf dem Spielerprofil die Forumsbeiträge der Marktwertdiskussion gelesen werden. Die Forumdiskussionen sind teilweise jedoch sehr lang und können zwischen den halbjährlichen Updates mehrere Dutzende, teilweise sehr ausführliche Beiträge enthalten.

Eine andere Möglichkeit, um nachvollziehen zu können, warum ein Spieler den Marktwert hat, gibt es nicht.

Dieser Herausforderung kann der Einsatz von künstlicher Intelligenz (KI) entgegentreten. Dafür kann eine KI die Zusammenfassung der Marktwertdiskussionen erstellen. Durch das hohe Beitragsvolumen ist dies ein interessantes Anwendungsszenario für KI. Mit dieser wäre eine Zusammenfassung der Diskussionsbeiträge bewältigbar. Innerhalb von Transfermarkt ist KI noch nicht Teil des Angebots. Prototypische KI-generierte Zusammenfassungen von Marktwertdiskussionen wurden zu Entwicklungszwecken bereits erstellt, jedoch sind diese Prototypen offline. Somit sind auch die relevanten Aspekte wie beispielsweise das Vertrauen in die Zusammenfassung und wie diese von den Nutzerinnen und Nutzern beurteilt werden, dem Transfermarkt Team unbekannt.

1.3 Zielsetzung

Durch diese Arbeit sollen zentrale Erkenntnisse über die Wahrnehmung der aktiven Nutzerinnen und Nutzer gegenüber KI-generierten Zusammenfassungen von Forumdiskussionen über Marktwerte gewonnen werden. Es soll evaluiert werden, wie die Nutzerinnen und Nutzer die Zusammenfassungen beurteilen und warum sie zu dieser Beurteilung kommen. Es soll auch erhoben werden, welchen Änderungsbedarf die Nutzerinnen und Nutzer gegenüber der prototypischen Zusammenfassung haben. Mit den gewonnenen Erkenntnissen soll die zukünftige Anwendung solcher KI-generierten Zusammenfassungen für die Nutzerinnen und Nutzer gewinnbringend implementiert werden können.

1.4 Fragestellungen

Auf Basis der geschilderten Zielsetzung werden folgende Fragestellungen formuliert:

«Wie bewerten aktive Nutzerinnen und Nutzer von Transfermarkt.de eine prototypisch KI-generierte Zusammenfassung einer Forumdiskussion über Marktwerte?»

«Welchen Änderungsbedarf haben aktive Nutzerinnen und Nutzer von Transfermarkt.de an die prototypisch KI-generierte Zusammenfassung von einer Forumdiskussion über Marktwerte?»

1.5 Aufbau der Arbeit

Diese Arbeit beginnt mit der Einleitung, die den Einstieg in die Thematik und die Arbeit bildet. Anschliessend wird die Thematik theoretisch verankert und damit eine Grundlage für die weitere Untersuchung gebildet. Danach wird auf die Methodik

eingegangen, das Untersuchungsinstrument beschrieben und die Datenauswertung erläutert. Im nächsten Schritt werden die Ergebnisse dargestellt und im anschließenden Kapitel diskutiert und interpretiert. Zum Abschluss wird ein Fazit gezogen und ein Ausblick gegeben.

1.6 Abgrenzung

Der Schwerpunkt dieser Arbeit bezieht sich darauf, wie die Nutzerinnen und Nutzer eine prototypisch KI-generierte Zusammenfassung einer Forumdiskussion über Marktwerte bewerten und wo sie möglichen Änderungsbedarf identifizieren. Die Arbeit hat nicht den Anspruch eine Übersicht über den derzeitigen Forschungsstand zur Qualität von KI-Zusammenfassungen zu geben. Auch sollen keine systematische Metriken zur Bewertung von Textzusammenfassungen (beispielsweise ROUGE (Ganesan, 2018)), zur Anwendung kommen. Des Weiteren werden keine ethischen und rechtlichen Aspekte der Nutzung von KI vom Autor gezielt thematisiert.

2 Theorie

Die im Theorieteil beschriebenen Inhalte dienen als Basis für das weitere theoriegeleitete Vorgehen in dieser Arbeit, welches im Methodenteil genauer dargelegt wird.

2.1 Human-Centered-Design-Prozess

Diese Arbeit stützt sich auf die Methoden der menschenzentrierten Gestaltungsprozesse (HCD-Prozess), welche den Menschen ins Zentrum der Forschung stellen (DIN EN ISO 9241-210, 2011).

Eine Möglichkeit, diese Gestaltungsprozesse umzusetzen, bietet der HCD-Prozess. Dieser ist ein menschenzentrierter, iterativer Prozess, der nicht nur die Nutzerinnen und Nutzer, sondern auch weitere Anspruchsgruppen, wie beispielsweise die Mitarbeitenden, in den Mittelpunkt der Gestaltung stellt. Das Produkt soll in verschiedenen Stadien immer wieder evaluiert werden (DIN EN ISO 9241-210, 2011). Diese verschiedenen Stadien finden sich auch in diesem Projekt wieder. In der untenstehenden Abbildung 1 ist der menschenzentrierte Gestaltungsprozess dargestellt. Diese Arbeit befindet sich demnach beim grün markierten letzten Schritt vor der ersten Iterationsschleife, dem Evaluieren einer Gestaltungslösung anhand der Anforderungen. Die Festlegung der Nutzungsanforderungen erfolgt in dieser Arbeit durch eine Literaturrecherche, welche die Grundlage für die weitere Untersuchung bildet.

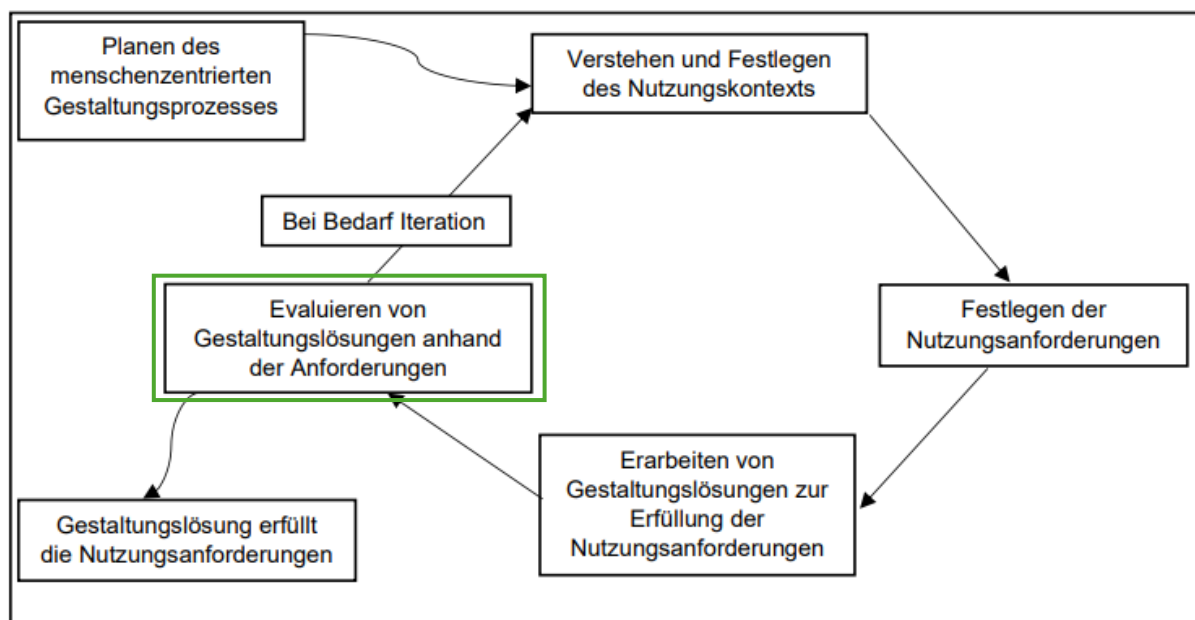


Abbildung 1. Menschenzentrierter Gestaltungsprozess, eigene Darstellung, nach DIN EN ISO 9241-210 (2011)

Gemäss Sarodnick und Brau (2016) kann eine frühzeitige Partizipation der zukünftigen Nutzerinnen und Nutzer zu einer höheren Systemqualität des Endprodukts und zu mehr Zufriedenheit und Akzeptanz führen. Auch Moser (2012) unterstreicht die

Wichtigkeit, dass die zukünftigen Nutzenden miteinbezogen werden und unter anderem ihre Bedürfnisse oder Ziele erforscht werden.

Diese Grundsätze gelten auch bei Anwendungen, die KI involvieren. Die Nutzerinnen und Nutzer sollen früh involviert sein, dadurch können die Anwendungen gezielt auf die Nutzenden abgestimmt werden. Probleme der Nutzerinnen und Nutzer sollen frühzeitig erkannt werden, sodass die Anwendung später einen tatsächlichen Mehrwert bieten kann. So können effektivere KI-gestützte Anwendungen entwickelt werden, welche die Nutzenden und ihre Fähigkeiten besser unterstützen („What is Human-Centered AI (HCAI)?“, 2025).

2.2 Künstliche Intelligenz und Zusammenfassungen

Die Kernaufgabe von KI ist, Aufgaben zu übernehmen, die bis anhin von Menschen ausgeführt wurden. KI-Systeme haben die Fähigkeit grosse Datenmengen zu verarbeiten, sie zu analysieren, Muster zu erkennen und daraus neue Inhalte zu generieren. Daraus ergibt sich ein breiter Einsatzbereich, von der Verarbeitung von menschlicher Sprache bis zur Erstellung von Bildern (Kreutzer, 2023).

KI-generierte Zusammenfassungen kommen beispielsweise bereits bei medizinischen und rechtlichen Texten zum Einsatz, oder auch bei Zusammenfassungen von Online-Dokumenten. Leserinnen und Lesern fällt es dabei oft schwer, zwischen KI-Zusammenfassungen und menschlichen Zusammenfassungen zu unterscheiden (Soni & Wade, 2023). Die Zusammenfassungen können das Bedürfnis nach tieferem Eintauchen in einen Quelltext nicht immer ausschliessen. Dies insbesondere bei komplexen und langen Quelltexten oder wenn die Zusammenfassungen sogenannte Halluzinationen enthalten (Kambhamettu, Flores & Head, 2024). Dennoch liefern Textzusammenfassungen von Chat-GPT gute Ergebnisse bezüglich der Kohärenz und der semantischen Übereinstimmung (Arnold & Hörauf, 2024). Auch Nutzerinnen und Nutzer betrachten KI-Zusammenfassungen als ein hilfreiches Werkzeug, jedoch sind zentrale, relevante Aspekte wie Transparenz, Kontrolle, Möglichkeit zur Anpassung und Vertrauen gegenüber KI zu beachten (Cheng, Smith-Renner, Zhang, Tetreault & Jaimes, 2022).

2.3 Relevante Aspekte

Durch eine Literaturrecherche konnten relevante und zentrale Aspekte bei durch KI zusammengefassten Inhalten identifiziert werden. Sie dienen als Grundlage für das weitere methodische Vorgehen, welches im Methodenteil detaillierter ausgeführt wird.

In den folgenden Abschnitten werden diese relevanten Aspekte ohne spezifische Reihenfolge geschildert.

2.3.1 Genauigkeit

Eine Nutzerstudie zu Zusammenfassungen von wissenschaftlichen Artikeln zeigte, dass Studentinnen und Studenten Zusammenfassungen besser bewerteten, welche die zentralsten Inhalte des Quelltextes enthielten (Yamamoto, Suzuki, Fukusato, Kataoka & Morishima, 2021). Die Wahrnehmung von Auslassungen oder Ungenauigkeiten hingegen führte bei den Teilnehmenden einer Untersuchung zu Zusammenfassungen von Datenschutzrichtlinien zu einer allgemeinen Skepsis gegenüber der Zusammenfassung (Bracamonte, Hidano, Tesfay & Kiyomoto, 2020). Zu bedenken ist, dass es sich bei beiden genannten Untersuchungen um Zusammenfassungen von komplexen Texten handelt und die Erkenntnisse nur bedingt übertragbar sind. Dennoch wird die Relevanz der Genauigkeit einer Zusammenfassung aufgezeigt und bestimmt die korrekte Wiedergabe des Quelltextes als Grundvoraussetzung einer guten Zusammenfassung. Dies konnten auch Ter Hoeve, Kiseleva & Rijke (2022) feststellen. Sie fanden bei den Nutzenden grosse Unterschiede bei den Anforderungen an den Detailgrad.

2.3.2 Umfang

Beim Nutzen einer Zusammenfassung sollte diese zu einer Zeit- und Aufwandsersparnis führen, ohne jedoch die Inhalte zu sehr zu vereinfachen. Vor allem in zeitkritischen Bereichen können Zusammenfassungen damit einen Mehrwert bieten (Stiglic et al., 2023). Andere Untersuchungen konnten zeigen, dass es unterschiedliche Bedürfnisse der Nutzenden bezüglich des Umfangs der Zusammenfassung gibt (Silvagnoli, Shepherd, Pritchett & Gardner, 2021). Auch (Ter Hoeve et al., 2022) weisen darauf hin, dass es unterschiedliche Bedürfnisse der Nutzenden an den Umfang geben kann und dies in die Erstellung einer Zusammenfassung mitbestimmen soll. Ebenfalls weisen sie darauf hin, dass es verschiedene Nutzungsanwendungen gibt und diese mitberücksichtigt werden sollen.

2.3.3 Vertrauen

Vertrauen in KI ist essenziell dafür, dass sie tatsächlich genutzt wird und erfolgreich angewendet werden kann („What is Human-Centered AI (HCAI)?“, 2025) Ter Hoeve et al. (2022) konnten feststellen, dass es beim Vertrauen auch auf die Qualität der Zusammenfassungen ankommt und weniger auf die Tatsache, dass sie von KI erstellt wurde. Zudem verwiesen sie auf die Qualität der Input-Daten, die ebenfalls berücksichtigt werden sollten. Eine hohe Transparenz und Nachvollziehbarkeit bei KI-Anwendungen kann nichtsdestotrotz zu einem höheren Vertrauen führen („What is Human-Centered AI (HCAI)?“, 2025).

Nutzende können bei KI-generierten Zusammenfassung auch Vertrauensbedenken haben, da sie sich zu sehr auf die Inhalte verlassen könnten, auch wenn diese nicht korrekt

sind. Ebenfalls kann sich eine hohe Komplexität des Ursprungstexts negativ auf das Vertrauen auswirken. Das Verständnis, wie eine Zusammenfassung zustande kam oder wie Abläufe bei der Entstehung sind, hat sich hingegen positiv auf das Vertrauen ausgewirkt (Cheng et al., 2022).

2.3.4 Transparenz und Kontrolle

Transparenz ist ein zentraler Bestandteil bei KI-generierten Inhalten. Es sollten klare Einblicke in die Anwendung gegeben werden, sodass die Nutzerinnen und Nutzer nachvollziehen können, wie die Inhalte entstanden sind. Dadurch können das Vertrauen in die KI und das Erlebnis bei der Nutzung gestärkt werden („What is Human-Centered AI (HCAI)?“, 2025).

Die Entstehung einer Zusammenfassung beeinflussen zu können und die Nachvollziehbarkeit der Entstehung sind ebenfalls wichtige Bestandteile einer gelungenen Zusammenfassung. So kann das Bedürfnis bestehen, die Entstehung zu beeinflussen, um die Qualität der Zusammenfassung zu erhöhen. Eine hohe Nachvollziehbarkeit und die Möglichkeit die Entstehung der Zusammenfassung beeinflussen zu können, können zudem das Vertrauen in die Zusammenfassung an sich stärken (Cheng et al., 2022).

2.3.5 Stil und Format

Nutzerinnen und Nutzer können gegenüber dem sprachlichen Stil unterschiedliche Bedürfnisse haben. So wurde beispielsweise bei einer zu komplexen Wortwahl das Verständnis schwieriger. Auch gab es Uneinigkeit darüber, wie sachlich und neutral eine Zusammenfassung verfasst werden sollte (Cheng et al., 2022). Die sprachliche Zusammenfassung kann zudem ergänzt werden mit strukturellen Elementen wie Hervorhebungen oder Grafiken. Solche Ergänzungen können zu einem besseren Erlebnis bei der Nutzung von Zusammenfassungen führen. Jedoch sind diese Bedürfnisse stark abhängig von den Nutzenden und können unterschiedlich ausgeprägt sein (Ter Hoeve et al., 2022)

2.3.6 Effizienz und Effektivität

Nutzerinnen und Nutzer können je nach Ziel und Kontext der Nutzung bei Zusammenfassungen unterschiedliche Bedürfnisse an die Effizienz haben. Wie gut die Effizienz ist, hat einen ausgeprägten Einfluss auf das Erlebnis bei der Nutzung (Cheng et al., 2022). Es ist wichtig, die Ziele zu kennen, welche die Nutzerinnen und Nutzer bei der Nutzung von Zusammenfassungen verfolgen und in welchem Kontext sie diese nutzen (Ter Hoeve et al., 2022).

Cheng et al. (2022) weisen zudem darauf hin, dass Effizienz im negativen Zusammenhang mit Transparenz und Kontrolle stehen kann. So kann eine hohe Effizienz zu geringen Kontrollmöglichkeiten führen oder die Transparenz zur Entstehung von KI-generierten Zusammenfassungen verringern. Zudem ist die wahrgenommene Effizienz wichtig für ein positives Nutzererleben.

3 Methoden

Dieses Kapitel erläutert das methodische Vorgehen von der Datenerhebung bis zur Datenauswertung. Als Erstes wird ein Überblick über den Untersuchungsplan gegeben. Anschliessend wird die Methode beschrieben und begründet, die Stichprobe sowie die Durchführung der Datenerhebung beschrieben und zum Schluss die Datenauswertung geschildert.

Das methodische Vorgehen stützt sich auf den qualitativen Forschungsprozess nach Flick (2019), bei dem Hypothesen und Theorien generiert und ein ganzheitlicher Ansatz verfolgt werden kann. Diese Herangehensweise ermöglicht ein theoriegeleitetes Vorgehen und eine möglichst inhaltlich reichhaltige Beantwortung der Fragestellung. Dafür eignen sich die qualitativen Forschungsmethoden für die Nutzerforschung wie beispielsweise Interviews (Moser, 2012).

3.1 Untersuchungsplan

In der nachfolgenden Abbildung 2 ist der Untersuchungsplan dargestellt und somit bildlich zusammengefasst.



Abbildung 2. Untersuchungsdesign (eigene Darstellung)

3.2 Halbstandardisierte Interviews

Für die Untersuchung der Forschungsfrage kommt ein selbst konstruierter halbstrukturierter Interviewleitfaden zur Anwendung. Der Interviewleitfaden wird auf der Basis der Erkenntnisse des Theorieteils konstruiert und orientiert sich an den beschriebenen relevanten Aspekten.

Interviews sind gut geeignet, um möglichst reichhaltige Informationen von einer eher kleineren Anzahl Personen zu sammeln. Für eine qualitative Forschung können halbstandardisierte oder unstandardisierte Interviews zum Einsatz kommen (Moser, 2012). Halbstandardisierte Interviews haben den Vorteil, dass mit dem Leitfaden sowohl

vorgegebene Fragen zum Untersuchungsgegenstand gestellt werden können und es dennoch Raum für spontane Fragen und Vertiefungen gibt. Somit können an passender Stelle gezielte Vertiefungsfragen gestellt werden und dennoch irrelevante Themen ausgeschlossen werden (Flick, 2019).

3.2.1 Erstellung des Leitfadens

Der Interviewleitfaden wird anhand der SPSS-Methode von Helfferich (2011) entworfen. In dieser Methode können Fragen für den Leitfaden in vier Schritten gefunden werden: Sammeln, prüfen, sortieren und subsumieren. Für die erste Sammlung von Fragen ist der zu untersuchende Forschungsgegenstand von entscheidender Bedeutung (Helfferich, 2011). Das Fundament für diese erste Sammlung bilden demnach die relevanten Aspekte, welche im Theorieteil beschrieben wurden. Die relevanten Aspekte waren Genauigkeit, Umfang, Vertrauen, Transparenz und Kontrolle, Stil und Format, Effizienz und Effektivität. Zu jedem dieser Aspekte werden mehrere erzählgenerierende und konkrete Fragen formuliert. Im Anschluss folgt eine kritische Prüfung der gebildeten Fragen in Hinblick auf deren Relevanz zur Beantwortung der Fragestellung, überflüssige werden gestrichen. Nach der Prüfung kommt die Sortierung der Fragen, dabei ist auf einen sinnvollen Zusammenhang und Ablauf zu achten, um eine logische und stimmige Struktur des Leitfadens sicherzustellen. Auch werden die Fragen anhand ihres Rangs in offene Erzählaufforderung, konkrete Nachfragen und Aufrechterhaltungsfragen eingeteilt. Abschliessend werden die Fragen subsumiert. Dabei erhält der Interviewleitfaden seine finale Form, mit vier Spalten, die verschiedene Funktionen innehaben. Die erste Spalte enthält die erzählgenerierende Frage. In der zweiten Spalte befindet sich eine Checkliste, welche die zentrale Thematik festhält und als Orientierungshilfe für die Interviewende Person dient. Die dritte Spalte enthält konkrete Fragen, welche gestellt werden, wenn nicht alle zentralen Themen angesprochen wurden. In der vierten und letzten Spalte sind Steuerungs- und Aufrechterhaltungsfragen festgehalten (Helfferich, 2011).

Der abschliessende Leitfaden wurde so strukturiert, dass den Interviewpersonen zuerst eine kurze Einführung gegeben wird und zwei Einstiegsfragen gestellt werden. Im Anschluss werden die Personen gebeten, einen Ausschnitt eines echten Diskussionsforums über einen Marktwert zu lesen. Danach können die Personen die entsprechende KI-generierte Zusammenfassung lesen. Im weiteren Verlauf des Interviews werden die Personen mithilfe des Forumsausschnitts und der Zusammenfassung mit den erzählgenerierenden Fragen, den konkreten Fragen und gegebenenfalls mit den Aufrechterhaltungs- und Steuerungsfragen befragt. Dabei wird mithilfe der definierten Checks sichergestellt, dass die jeweils zentrale Thematik behandelt wurde. Die Zeile, mit der ersten erzählgenerierenden Frage, die einen relevanten Aspekt behandelt, ist in der

untenstehenden Tabelle 1 beispielhaft dargestellt. Der gesamte Leitfaden ist im Anhang unter Kapitel C aufgeführt.

<u>Genauigkeit</u>	<u>Checks</u>	<u>Konkrete Fragen</u>	<u>Aufrechterhaltungs- und Steuerungsfragen</u>
Finden Sie, dass die Zusammenfassung die Diskussion genau abbildet?	<ul style="list-style-type: none"> • Beurteilung der Genauigkeit • Änderungsbedarf 	<ul style="list-style-type: none"> • Was finden Sie ist gut / schlecht abgebildet? • Warum finden Sie, dass das gut / schlecht abgebildet ist? • Was würden Sie ändern, um die Forumsbeiträge besser abzubilden? 	<ul style="list-style-type: none"> • Warum wirkt das so auf Sie? • Wie haben Sie es verstanden?

Tabelle 1. Ausschnitt aus Interviewleitfaden

3.2.2 Auswahl der Forumsbeiträge

Um den erwähnten Ausschnitt aus dem Diskussionsforum zu bestimmen, wurden die Foren von verschiedenen deutschsprachigen Spielern stichprobenartig nach einer passenden Diskussion durchsucht. Es wurde darauf geachtet, dass es sich um eine aktuelle Diskussion handelt, sodass bei den Interviewteilnehmenden keine Verwirrung oder Ablenkung durch mögliche unzeitgemässe Argumente oder veraltete Darstellungen entstehen könnte. Zudem wurde bei der Auswahl sichergestellt, dass die einzelnen Diskussionsbeiträge zusammenhängen und nicht durch eine mehrwöchige Pause unterbrochen waren. Ebenfalls wurde berücksichtigt, dass der Forumsausschnitt nicht zu lange und innert 5 Minuten lesbar ist. Der gewählte Ausschnitt ist im Anhang unter Kapitel A ersichtlich.

3.2.3 Erstellung der prototypischen Zusammenfassung

Die Zusammenfassung des ausgewählten Ausschnitts aus dem Diskussionsforum wurde auf Basis des bestehenden Prototyps erstellt. Dafür wurde dem KI-Modell ChatGPT-4o der Prompt «Fasse mir folgende Forumdiskussion über den Marktwert eines Fussballspielers kurz und in reiner Textform zusammen:» angegeben und der gewählte Ausschnitt eingefügt. Die daraus entstandene Zusammenfassung ist im Anhang unter Kapitel B festgehalten.

3.2.4 Pre-Test Interviewleitfaden

Vor dem ersten Interview wurde ein Pre-Test durchgeführt. Dabei hat sich herausgestellt, dass der gewählte Ausschnitt des Diskussionsforums zu viele Beiträge enthält und es daher schwer ist, sich diese zu merken. Daraus resultierte ein erschwerter Vergleich zwischen dem Diskussionsforum und der Zusammenfassung. Um dies zu verbessern, wurden die berücksichtigten Beiträge im Forum von 16 auf 11 gekürzt. Dies 11 Beiträge wurden ohne grosse zeitliche Abstände verfasst und waren Bestandteil einer

zusammenhängenden Diskussion. Ebenfalls wurden einige Fragen in der Formulierung leicht angepasst, um die Verständlichkeit und den Interviewfluss zu verbessern.

3.2.5 Beschreibung der Stichprobe

Für die Untersuchung wurden 12 aktive deutschsprachige Nutzerinnen und Nutzer von Transfermarkt.de rekrutiert. Sämtliche Untersuchungspersonen sind regelmässige Nutzende von Transfermarkt.de und haben ein grundsätzliches Verständnis über Marktwerte. Die Interviewpersonen waren zwischen 24 und 34 Jahre alt, davon waren 3 Personen weiblich und 9 männlich. Von den interviewten Nutzerinnen und Nutzer gaben 4 Personen an, dass sie Transfermarkt.de monatlich oder mehrmals monatlich besuchen. 7 Personen gaben an, dass sie Website wöchentlich oder mehrmals wöchentlich besuchen. 1 Person gab an, dass sie täglich Transfermarkt besucht.

3.2.6 Datenerhebung

Von den insgesamt 12 durchgeführten Interviews wurden 7 vor Ort und 5 remote durchgeführt. Den Interviewpersonen wurde im Laufe des Interviews aufgefordert den Ausschnitt des Diskussionsforums und der entsprechenden Zusammenfassung zu lesen. Sie hatten dabei die Möglichkeit Notizen oder Markierungen zu machen, dies war jedoch freiwillig und den teilnehmenden Personen überlassen. Die Interviews dauerten inklusive Einführung, Lesezeit der erwähnten Texte und Abschluss durchschnittlich 30 Minuten. Das kürzeste Interview dauerte 23 Minuten und das Längste 38 Minuten.

3.3 Datenauswertung

Die Transkription und Auswertung der qualitativen Daten wurde mit der Daten- und Textanalyse Software MAXQDA 2024 durchgeführt. Die aufgezeichneten Audiodateien der Interviews wurden vorab mit Turboscribe vortranskribiert. Dabei handelt es sich um ein KI gestützten Transkriptionstool. Im Anschluss wurden die Dateien in MAXQDA 2024 eingefügt. Dort wurden sie detailliert transkribiert und die Kodierung sowie die Auswertung durchgeführt. Alle Transkripte sind Anhang unter Kapitel E aufgelistet.

3.3.1 Transkription

Die detaillierte Transkription erfolgte nach den Regeln von Rädiker und Kuckartz (2019). Durch die Einhaltung dieser Regeln wird die Erstellung des Kategoriensystems erleichtert. Diese sind verständlich formuliert und für verschiedene Interviewformer geeignet. Nachfolgend werden die Regeln in der Tabelle 2 zusammengefasst.

1. Die Transkription erfolgt wortwörtlich, somit nicht lautsprachlich und nicht zusammenfassend. Dialekte werden nicht transkribiert, sondern weitgehend genau ins Hochdeutsche übersetzt.
2. Um Rückschlüsse auf die Interviewperson zu verhindern, werden Angaben, die ein solcher Rückschluss erlauben, anonymisiert und wenn dies nicht geht, ganz weggelassen.
3. Die Sprache sowie die Interpunktion werden geglättet und somit dem Schriftdeutsch angenähert. Die Wortstellung und die Artikel werden bleiben unverändert, auch wenn darin Fehler enthalten sind.
4. Längere Pausen werden mit in Klammer gesetzten Punkten gekennzeichnet.
5. Irrelevante Aussagen und Wiederholungen werden nicht transkribiert.
6. Besondere Betonungen werden durch Unterstiche markiert.
7. Kurze Einwürfe, wie beispielsweise «okay», der anderen Person werden in Klammer gesetzt und in den Sprechbeitrag eingefügt.
8. Ebenfalls werden Lautäusserungen beider Personen in Klammern erwähnt. Beispielsweis, wenn eine Person lacht, wird dies in Klammer als (lacht) festgehalten.

Tabelle 2. Transkriptionsregeln (in Anlehnung an Rädiker und Kuckartz, 2019, S. 44-45)

3.3.2 Kategoriensystem

Die weitere Verarbeitung der Transkripte erfolgt mit der inhaltlich strukturierenden Analyse nach Kuckartz (2018). Anhand dieser Methode werden bei der Kodierung die Hauptkategorien und einige Subkategorien deduktiv aus dem Interviewleitfaden abgeleitet. So wurden die Hauptkategorien Genauigkeit, Umfang, Vertrauen, Transparenz und Kontrolle, Stil und Format, Effizienz und Effektivität und weitere Anwendungen aus dem Leitfaden abgeleitet. Abgesehen von der Subkategorie Grafiken und Hervorhebungen wurden sämtliche Subkategorien induktiv aus den Daten abgeleitet. Die Subkategorie Grafiken und Hervorhebungen wurde deduktiv aus dem Leitfaden gebildet, da eine konkrete Frage dazu gestellt wurde. Es wurde erfragt, ob die Nutzerinnen und Nutzer solche Grafiken und Hervorhebungen begrüßen würden. Diese Frage wiederum basierte auf dem spezifischen Hinweis, welcher im Theorieteil unter Kapitel 2.3.5 beschrieben wurde.

Bei der Kodierung wurden zuerst die Hauptkategorien deduktiv kodiert. Anschliessend konnten Textabschnitte innerhalb derselben Hauptkategorie differenziert betrachtet und induktiv unterteilt werden. Das Kategoriensystem wird in der folgenden Abbildung 3 dargestellt. Die Beschreibungen der Hauptkategorien sind im Anhang im Kapitel D genauer beschrieben. Die Beschreibungen der Subkategorien gehen aus dem Ergebnisteil hervor.

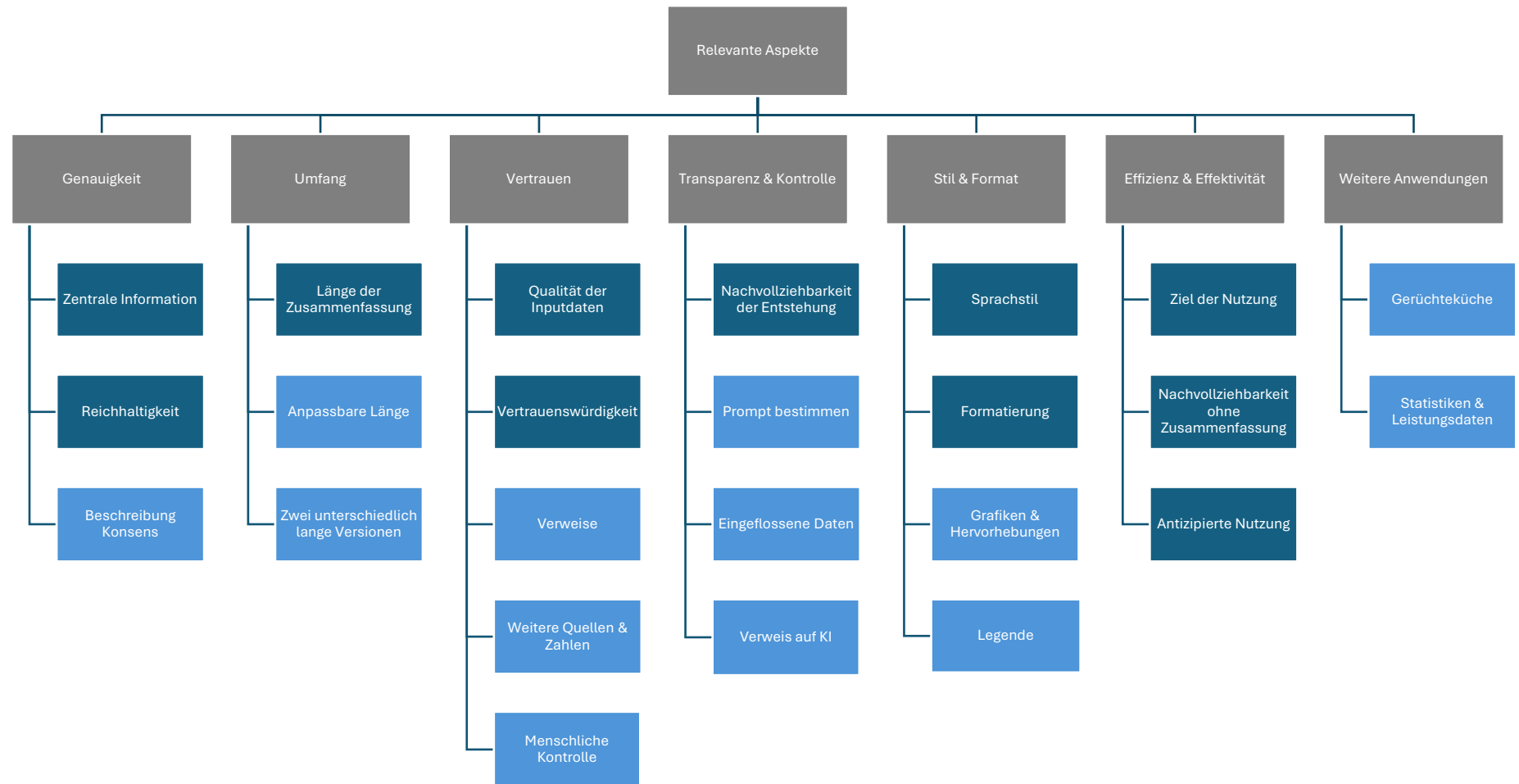


Abbildung 3. Kategoriensystem (eigene Darstellung) Dunkelblau stellt die Bewertung dar, Hellblau stellt den Änderungsbedarf dar

3.3.3 Auswertung

Die Ergebnisse und somit die Beantwortung der Fragestellung leiten sich aus dem gebildeten Kategoriensystem ab. Nach Kuckartz und Rädiker (2024) stellt das Kategoriensystem an sich ein wertvolles Ergebnis der analytischen Arbeit dar. Weiter sind die Daten durch das mehrstufige Kategoriensystem sehr fein gegliedert und durch den systematischen Zuordnungsprozess wohlstrukturiert. Somit ist eine detaillierte Beantwortung der Forschungsfragen möglich (Kuckartz und Rädiker, 2024).

Die abschliessende Auswertung erfolgt entlang der Hauptkategorien. Die Forschungsergebnisse werden anhand des Kategoriensystems, beziehungsweise anhand der Hauptkategorien gegliedert.

Im Ergebnisteil werden die inhaltlich forschungsrelevanten Gesichtspunkte sämtlicher codierten Textstellen entlang der Haupt- und Subkategorien nah am transkribierten Text, aber mit eigenen Worten, beschrieben. Die Forschungsergebnisse werden zudem durch prototypische Zitate aus dem Datenmaterial unterstützt (Kuckartz, 2018).

4 Ergebnisse

In den folgenden Kapiteln werden die Ergebnisse anhand des Verlaufs des Kategoriensystems dargelegt. Zum allgemeinen Verständnis wird innerhalb von jedem Unterkapitel zuerst auf die erste Fragestellung «Wie bewerten aktive Nutzerinnen und Nutzer von Transfermarkt.de eine prototypisch KI-generierte Zusammenfassung einer Forumdiskussion über Marktwerte?» und anschliessend auf die zweite Fragestellung «Welchen Änderungsbedarf haben aktive Nutzerinnen und Nutzer von Transfermarkt.de an die prototypisch KI-generierte Zusammenfassung von einer Forumdiskussion über Marktwerte?» eingegangen. Dafür wird die Hauptkategorie beschrieben und anschliessend die Subkategorien mit passenden prototypischen Zitaten erläutert. Wie bereits erwähnt, sind alle Transkripte im Anhang unter Kapitel E aufgelistet.

4.1 Genauigkeit

Die Hauptkategorie Genauigkeit umfasst zwei Subkategorien, welche den Inhalt der Zusammenfassung gegenüber der ursprünglichen Diskussion beurteilen und eine Subkategorie, die den Änderungsbedarf benennt.

4.1.1 Bewertung

In diesem Abschnitt werden die Meinungsbilder der Interviewteilnehmenden gegenüber der Genauigkeit zusammenfasst. Es liessen sich zwei Subkategorien zur Bewertung der Genauigkeit feststellen und ein Änderungsbedarf wurde genannt.

Zentrale Information

Alle interviewten Personen empfanden, dass die Zusammenfassung die zentralsten und wichtigsten Informationen ausgewogen darstellt. Wenige Personen haben einzelne Punkte festgestellt, die sie als schlecht abgebildet empfunden haben. Eine Person hat es wie folgt beschrieben:

Ich finde die Zusammenfassung eigentlich, nachdem ich natürlich auch die Kommentare gelesen habe, passt schon sehr gut zusammen. (9. Interview, Pos. 10)

Zwei Personen fanden, dass die Aussage der Forumsmitglieder, welche einen tieferen Marktwert von unter 25 Millionen vorschlugen, überrepräsentiert war in der Zusammenfassung. Dennoch urteilten beide Personen positiv über den Inhalt der Zusammenfassung. Eine Person hat es folgendermassen beschrieben:

Ich denke, sie wird ziemlich gut abgebildet. Worauf man achten muss, ist, dass einige Stimmen nicht ganz repräsentiert sind in der Anzahl. Das heißt, es wird von einzelnen

Stimmen gesprochen, die Konrad Laimer in diesem Beispiel mit einem Marktwert von unter 25 Millionen zählen. Aber in der realen Quelle, also in der Diskussion, ist lediglich ein Nutzer davon überzeugt. Und da es von einzelnen Stimmen in der Mehrzahl hier jetzt in der Zusammenfassung steht, könnte dieser Wert ein bisschen anders interpretiert werden.

An der Zusammenfassung wurde zudem von vielen geschätzt, dass nicht nur die zentralen Informationen vorkamen, sondern auch die Argumente ausgewogen präsentiert wurden. Eine Person hat es wie folgt beschrieben:

Nein, also ich finde, es ist zusammengefasst. Ich habe das Gefühl, dass sie eigentlich schaut, welchen gemeinsamen Nenner es gibt, welche Stimmen oder Aussagen am häufigsten vorkommen und gibt das dann wieder. Nichtsdestotrotz weist die Zusammenfassung auch darauf hin, dass es Stimmen gibt, die diesen Marktwert deutlich höher oder deutlich darunter sehen. Aber sie fasst wie den Konsens zusammen. (2. Interview, Pos. 16)

Reichhaltigkeit

Zur Reichhaltigkeit gab es unterschiedliche Stimmen, obwohl sie von der Mehrheit als positiv bewertet wurde. Eine Person fand, dass noch mehr detaillierte Argumente aufgenommen werden könnten. Sie hat es folgendermassen beschrieben:

Meiner Meinung nach ist die Diskussion eher sehr grob umschrieben. Es gibt einen groben Überblick über die Diskussionen, jedoch sieht man nicht die Details oder Einzelheiten, die die Nutzer auch beschreiben. (11. Interview, Pos. 14)

Eine andere Person empfand die Reichhaltigkeit als ausreichend und würde nicht mehr Informationen in der Zusammenfassung erwähnen. Die Person hat es wie folgt beschrieben:

Ich finde, dass die Zusammenfassung besonders gut auf die verschiedenen Meinungen eingeht und sie aber trotzdem zusammenfasst. (10. Interview, Pos. 16)

Eine differenziertere Aussage zum Inhalt wurde von einer weiteren Person wie folgt getroffen:

Meiner Meinung nach hat die Zusammenfassung die wichtigsten Punkte zusammengefasst. Details, wie dass sie nicht das CL-Finale gespielt haben, sind außen vor gelassen. Was ich wiederum sehr gut finde. (7. Interview, Pos. 14)

4.1.2 Änderungsbedarf

Der Änderungsbedarf bei der Genauigkeit viel gering aus. Es gab ein einzelnes Votum, die einen Bedarf bekundeten.

Beschreibung Konsens

Eine Person würde es bereichernd finden, wenn die KI den Konsens in der Forumdiskussion einschätzt. Sie hat es wie folgt beschrieben:

Um die Zusammenfassung besser nachvollziehen zu können, wäre es noch interessant, wenn KI den Konsens noch beschreiben würde, ob sich die Personen einig sind oder nicht (lacht). Ich denke, das wäre noch sehr hilfreich für die Einschätzung. (7. Interview, Pos. 57)

4.2 Umfang

Die Hauptkategorie Umfang beinhaltet Subkategorien bei denen die Länge der Zusammenfassung bewertet wurde und es konnten zwei Subkategorien zum Änderungsbedarf festgestellt werden.

4.2.1 Bewertung

Länge der Zusammenfassung

Bei der Länge der Zusammenfassung herrscht ein einheitliches Meinungsbild. Die Mehrheit der zwölf interviewten Personen empfand diese als passend und gut. Eine interviewte Person schätzte es wie folgt ein:

Wie bereits erwähnt, war für mich persönlich der Umfang sehr gut. Es war nicht zu knapp. Es hat nichts Wesentliches gefehlt und war nicht lückenhaft. Es war ausführlich, aber nicht zu lang, damit man sich den Inhalt gut merken konnte. (1. Interview, Pos. 27)

Im Zusammenhang mit der angemessenen Länge erwähnten einige Personen, dass die Zusammenfassung so zum Lesen und zur Nutzung animiert. So sollte eine Zusammenfassung nicht zu lange sein, sodass der Aufwand, diese zu lesen nicht zu gross ist. Eine Person hat es so beschrieben:

Ich finde den Umfang passend. Es dürfte nicht mehr sein, da es sonst nicht animiert zum Lesen und für einen kurzen Einblick reicht dies völlig aus. (7. Interview, Pos. 19)

Eine interviewte Person fand, dass die Länge der Zusammenfassung im Verhältnis zum Forum zu kurz war und mehr hätte zusammengefasst werden können. Die Person hat es folgendermassen beschrieben:

Ja, ich finde, für das es sieben Nutzer ausgewertet wurden, nur drei kleine Abschnitte sind etwas zu mager. Da ich auch alle Nutzer durchgelesen habe, habe ich das Gefühl, es könnte viel mehr zusammengefasst werden. (4. Interview, Pos. 80)

4.2.2 Änderungsbedarf

Beim Änderungsbedarf liessen sich zwei Subkategorien feststellen.

Anpassbare Länge

Zwei Personen machten den Vorschlag, dass die Länge der Zusammenfassung anhand der Einigkeit und der Länge der Ursprungsdiskussion angepasst werden kann. Somit könnte sichergestellt werden, dass die Zusammenfassung nicht unnötig lang oder zu kurz für die Anzahl unterschiedlicher Meinungen ist. Eine Person beschrieb es so:

Nicht nur anhand der Länge der Diskussion, sondern auch anhand der verschiedenen Themen der Diskussion. Wenn viele Nutzer der gleichen Meinung sind, kann man die Zusammenfassung sicher kürzen. Wenn es mehrere Diskussionen und Themen gibt, sollte man das schon ein bisschen länger machen. (11. Interview, Pos. 24)

Zwei unterschiedlich lange Versionen

Zwei Interviewte haben vorgeschlagen, zwei unterschiedlich lange Versionen zur Verfügung zu stellen. Somit könnten die Nutzerinnen und Nutzer selbst entscheiden, welche Version sie lesen wollen. Eine Person hat es wie folgt beschrieben:

Das kommt darauf an. Wenn man es detaillierter mag, dann könnte es schon länger sein. Man könnte auch eine kurze Zusammenfassung und dann nach unten einen Punkt, wo eine detaillierte Zusammenfassung ist, aufschalten, damit die Leser oder Leserinnen das selber entscheiden können, welche Version sie gerne lesen würden. (6. Interview, Pos. 25)

Dieser Vorschlag steht in Einklang mit einer Aussage einer weiteren Interviewperson, die zwar keinen Änderungsbedarf bekundete, jedoch darauf hinwies, dass es unterschiedliche Bedürfnisse gegenüber der Länge geben könnte. Die Person hat es wie folgt beschrieben:

Wenn das Ziel ist, dass man einen kurzen Überblick bekommt, dann dürfte es noch ein bisschen knackiger sein, noch ein bisschen weniger. Es kommt auf den Nutzer oder die Nutzerin an. Wenn eher das Interesse besteht, die ganze Diskussion noch ein

bisschen mehr abgebildet zu sehen, dann dürfte es noch ein, zwei Sätze länger sein.
(2. Interview, Pos. 24)

4.3 Vertrauen

Die Hauptkategorie Vertrauen beinhaltet zwei Subkategorien zur Bewertung und drei Subkategorien zum Änderungsbedarf. Bei der Bewertung ging es um die Qualität der Inputdaten und um die Vertrauenswürdigkeit von KI.

4.3.1 Bewertung

Qualität der Inputdaten

Sieben Interviewpersonen hinterfragen die Qualität der Inputdaten, beziehungsweise die Qualität der Kommentare im Forum. So gaben die Personen zu bedenken, dass die KI-Zusammenfassung nur so gut sein kann, wie die Daten, mit denen sie gefüttert wurde. Eine Person hat es wie folgt ausgedrückt:

Gut, die KI ist immer nur so gut wie das, was reinkommt. Also wie gesagt, ich glaube, die Zusammenfassung ist so gut. Ich kenne die Zusammenfassung von KI, ich finde die gut. Das funktioniert definitiv. Kommt halt, wie gesagt, stark darauf an, wie die Diskussion im Hintergrund verlief, wie sich die Leute ausdrücken. Von dem her, ich weiß, dass es funktioniert. Ich bin mir überzeugt, dass es funktionieren kann. Die Importdaten müssen einfach gut genug sein. (9. Interview, Pos. 25)

Eine Person stellte fest, dass es vermutlich schwierig wäre, die inhaltliche Qualität der Forumsbeiträge oder die Expertise der Forumsmitglieder zu beurteilen. Sie hat es so beschrieben:

Was auch vielleicht clever wäre, ich weiß nicht, wie das geht, den Hintergrund dieser Menschen irgendwie zu beleuchten. Wer diskutiert dort? Welche Aussagekraft kann man diesen Nutzern zuschreiben? Haben sie schon viel kommentiert? Sind sie schon lange dabei? Sind sie heute das erste Mal und geben irgendeine Einschätzung? Haben Sie überhaupt etwas mit Fußball zu tun? Das ist halt ungewiss. Wer sind die Nutzer? Welche Erfahrung, welches Know-how haben die, um diese Aussagen zu tätigen und damit auch ihre Aussagen glaubwürdig sind? (8. Interview, Pos. 20)

Eine weitere Person machte eine ergänzende Feststellung und wies darauf hin, dass auch die Art und Weise, wie man die Daten der KI eingibt, zu Qualitätsunterschieden führen könnte. Die Person hat es wie folgt erläutert:

Es kommt aber auch ein bisschen darauf an, wie man diese KI füttert. Ich denke, wenn man das einheitlich macht und immer auf die gleiche Weise, dann kann das sehr gut werden und kann auch sehr vielversprechend sein. Wenn man dann aber die Diskrepanzen hat, wie man die KI füttert, kann es sein, dass dann die Qualität sehr unterschiedlich ist. (6. Interview, Pos. 27)

Vertrauenswürdigkeit

Die Vertrauenswürdigkeit beinhaltet differenzierte Aussagen dazu, ob und warum die Interviewpersonen der KI-generierten Zusammenfassung vertrauen oder nicht. Die Meinungen waren unterschiedlich, so gaben neun Personen an der Zusammenfassung grundsätzlich zu vertrauen. Jedoch gab es unterschiedliche Vorbehalte, weshalb der Zusammenfassung nicht blind vertraut werden sollte. Eine Person drückte es folgendermassen aus:

Schlussendlich ist es eine Zusammenfassung, welche generiert wird, also ich vertraue diesem jetzt. Es kann immer sein, dass mal ein Fehler eingespielt wird, aber das wäre vielleicht bei einer menschlichen Zusammenfassung nicht anders. Darum finde ich das wirklich, also vertraue ich dem. (12. Interview, Pos. 30)

Eine andere Person drückte auch grundsätzliches Vertrauen mit Vorbehalten aus:

Ja, ich denke, KI-generierte Zusammenfassungen sind vertrauenswürdig, aber nie ohne das Originaldokument gegenzulesen. Es ist gut, um sich einen kurzen Überblick zu verschaffen oder um einen Einstieg zu bekommen in ein Thema, aber in der Originalquelle steckt dann immer noch mehr Information drin. Was die KI gut kann, ist das Runterbrechen und relativ neutral darzustellen. (8. Interview, Pos. 28)

Drei Personen äusserten ein gewisses Misstrauen gegenüber den Fähigkeiten von KI. Eine interviewte Person hat es wie folgt beschrieben:

Ich habe das Gefühl, wir sind immer noch recht am Anfang, was KI und ChatGPT angeht. Darum würde ich sagen, das Vertrauen meinerseits ist da nicht so hoch. Ich hätte keine Erwartungen, oder ich hätte gar nicht das Vertrauen, eine Arbeit mit dem schreiben zu lassen, da ich einfach das Gefühl habe, da gibt es noch viel Verbesserungspotenzial. Darum würde ich mich nicht auf das verlassen. (4. Interview, Pos. 34)

Von allen Personen begründeten neun ihre Einschätzung mit den Erfahrungswerten gegenüber KI, die sie bereits im Umgang mit KI gesammelt haben. Eine Person hat ihr positives Vertrauen so erklärt:

Weil ich schon öfters mit KI zusammengearbeitet habe und es auch in meinem Alltag brauche. Und ich daher einen guten Bezug zur Arbeit mit KI habe. (10. Interview, Pos. 33)

Hingegen erklärte eine andere Person ihr geringes Vertrauen ebenfalls durch einen gewissen Erfahrungswert:

Bei KI bin ich momentan noch ein bisschen vorsichtig, weil sie teilweise noch sehr fehleranfällig sind. (6. Interview, Pos. 27)

4.3.2 Änderungsbedarf

Beim Änderungsbedarf konnten drei Subkategorien festgestellt werden. Die meisten Stimmen wurden zum ersten Änderungsvorschlag gegeben, den Verweisen.

Verweise

Sieben interviewte Personen haben angegeben, dass sie Verweise auf das Diskussionsforum bereichernd finden würden und dass diese das Vertrauen stärken könnten. Es gab verschiedenste Vorschläge zu möglichen Verweisen, von Quellenangaben, die zur spezifischen Aussage im Forum führen, bis zu Verweisen darauf, von welchem Nutzer die Argumente kamen. Mit den Verweisen besteht die Möglichkeit, sich in spezifischen Abschnitten vertieft einzulesen oder die unterschiedlichen Argumente anhand der Nutzerin oder des Nutzers, die sie verfasst hat, einzuordnen. Eine Person hat es wie folgt beschrieben:

Um das Vertrauen zu erhöhen, könnte man aus meiner Sicht noch jeweils Verweise anfügen im Text, wo anschließend direkt zu der Diskussion geführt wird, um, wenn man noch Fragen hat oder nicht ganz sicher ist, genau nachlesen zu können. (7. Interview, Pos. 27)

Weitere Quellen und Zahlen

Von den Interviewten wünschten sich vier Personen, dass in die Zusammenfassung noch weitere Quellen und Zahlen einfließen, da damit das Vertrauen gestärkt werden könnte. Vorschläge kamen zu statistischen Werten oder externen Quellen, die einbezogen werden könnten. Auch der Vorschlag, dass sich die Zusammenfassung mehr auf die genannten Zahlen im Diskussionsforum bezieht, wurde als Vorschlag angebracht. Eine Person hat es wie folgt eingeordnet:

Für mich sind die Leistungsdaten aussagekräftig, weil das sind Fakten. Die Kommentare sind auch persönlicher Natur. (7. Interview, Pos. 35)

Eine andere Person hat auf anderweitige Quellen verwiesen, die einfließen könnten und hat es so beschrieben:

Eventuell könnte man auch zu den Kommentaren vielleicht noch andere Quellen dazunehmen. Eine Einschätzung, wie es in den sozialen Medien oder in anderen Foren widerspiegelt und eine einheitliche Erklärung abgeben. (6. Interview, Pos. 31)

Menschliche Kontrolle

Eine einzelne Stimme schlug vor, dass das Vertrauen durch menschliche Kontrolle verbessert werden könnte. Die Interviewperson hat es wie folgt beschrieben:

Was da helfen könnte. Vielleicht trotzdem, dass dann jemand noch auf die KI-gestützte Zusammenfassung, also dass noch ein Mensch darüber schaut, dass da wie noch eine Kontrollfunktion bei einem Menschen bleibt. Das würde wahrscheinlich schon mein Gefühl verbessern. (2. Interview, Pos. 32)

4.4 Transparenz und Kontrolle

Zu Transparenz und Kontrolle liess sich ein differenziertes Meinungsbild finden. So wurde die Anforderung an die Transparenz über die Entstehung differenziert beurteilt. Bei der Bewertung gab es eine Subkategorie, jedoch mit unterschiedlichen Ausprägungen. Beim Änderungsbedarf konnten drei Subkategorien festgestellt werden, so gab es unterschiedliche Meinungen dazu, inwiefern Kontrolle über die Zusammenfassungen gewünscht ist.

4.4.1 Bewertung

Nachvollziehbarkeit der Entstehung

Es gab unterschiedliche Meinungen dazu, wie wichtig die Nachvollziehbarkeit der Entstehung der Zusammenfassung ist. Sieben Personen bekundeten Desinteresse daran, wie die Zusammenfassung entstanden ist. Eine Person empfand es als unwichtig. Als Gründe dafür wurden beispielsweise die fehlende Tragweite der Thematik genannt oder gesagt, dass es auf den Inhalt ankommt und diesem grundsätzlich vertraut werden kann. Eine Person hat es wie folgt beschrieben:

Ich glaube, ich habe schon ein sehr großes Vertrauen in diese zukünftige künstliche Intelligenz, in unterschiedliche Tools. Und jetzt ich vertraue eigentlich wirklich auf diese Zusammenfassungen. Jetzt in Ihrem Beispiel. (1. Interview, Pos. 43)

Eine andere Person hat es folgend beschrieben:

Schlussendlich ist es mir eigentlich egal, wie die Zusammenfassung entstanden ist, solange ich die nötigen Zahlen und Fakten lesen kann und für mich verständlich ist,

wie dieser Marktwert zustande gekommen ist, ist mir eigentlich egal, wie diese Zusammenfassung geschrieben wurde. (4. Interview, Pos. 40)

Nebst den genannten Beurteilungen gab es vier Personen, welche die Nachvollziehbarkeit als wichtig oder interessant beurteilten. Jedoch äusserten zwei davon auch gewisse Vorbehalte, da sie es nur bedingt so beurteilten. Eine Person hat es wie folgt geschildert:

Meiner Meinung nach ist es schon noch interessant, wenn man weiss, wie die Zusammenfassung jetzt zustande kommt, dass es einen kleinen Überblick gibt. Für mich persönlich ist das nicht der springende Punkt. (11. Interview, Pos. 34)

Zwei Personen wiesen zusätzlich darauf hin, dass es für sie interessant und wichtig sei, den Prompt einsehen zu können, der verwendet wurde.

4.4.2 Änderungsbedarf

Prompt bestimmen

Den Prompt, welcher zur Zusammenfassung führt, zu bestimmen, fand eine Person eine interessante Änderung. Sie würde es spannend finden, die Schwerpunkte der Zusammenfassung setzen zu können. Jedoch haben sechs Interviewte gesagt, dass sie das nicht als nötig empfinden und bekundeten kein konkretes Interesse daran. Eine Person hat es folgendermassen beschrieben:

Ich denke, ich würde das gerne voreingestellt haben. That`s it! ich konsumiere, das muss schnell gehen, das ist eine Zusammenfassung, darum bin ich dafür, keinen Einfluss darauf zu nehmen. (9. Interview, Pos. 31)

Eine andere Person, die es zwar grundsätzlich interessant fand, aber dennoch Ablehnung bekundete, hat es wie folgt beschrieben:

Wäre sicher sehr interessant, der KI-Inputs zu liefern. Aber im Endeffekt geht es ja eigentlich um ein Thema, dass ich einfach eine Information schnellstmöglich aufsaugen möchte und dann mit dem eigentlich weiterfahren kann. (3. Interview, Pos. 42)

Eingeflossene Daten

Fünf interviewte Personen haben ausgesagt, dass sie es wichtig finden, dass die eingeflossenen Daten grundsätzlich referenziert werden. Es sollte nachvollziehbar sein, von wo die Daten stammen, die zusammengefasst wurden und wie diese aufzufinden sind. Auch wurden Angaben zu den Daten an sich, also beispielsweise wie viele Forumsbeiträge

zusammengefasst wurden, als hilfreich beschrieben. Eine Person hat es wie folgt geschildert:

Für mich persönlich gibt es eine grössere Sicherheit, wenn ich diese Zusammenfassung nehme. Wenn ich auch weiss, wie genau die zustande kam, was für Daten wurden genommen für die Zusammenfassung, gibt mir persönlich einfach noch ein bisschen grössere Sicherheit. (11. Interview, Pos. 38)

Verweis auf KI

Ein weiterer Änderungsbedarf wurde von drei Personen geäussert. Sie stufen es als wichtig ein, dass bei der Zusammenfassung transparent auf die Anwendung von KI hingewiesen wird. Eine Person hat es folgendermassen beschrieben:

Ich bin starker Fan davon, dass wenn man die KI verwendet, dass man das auch hinschreibt, dass es KI generierte Texte sind. Ich glaube zumindest in der heutigen Verfassung der KI können immer noch sehr viele Fehler passieren, dass man sich da auch ein bisschen schützt und das da hinschreibt, dass es ein KI generierter Text ist. Ich denke, das wäre schon wünschenswert. (9. Interview, Pos. 29)

4.5 Stil und Format

Beim Stil und Format geht es darum, wie die Zusammenfassung dargestellt wird und welche darstellerischen Elemente zum Einsatz kommen.

4.5.1 Bewertung

Aus den Daten konnten zwei Subkategorien abgeleitet werden. Sie beschreiben einerseits den sprachlichen Stil und andererseits die Formatierung der Zusammenfassung.

Sprachstil

Der Mehrheit gefiel der sprachliche Stil in der Zusammenfassung. Dieser wurde oft als angemessen und einfach verständliche beschrieben. Eine Person formulierte es wie folgt:

Man versteht sehr leicht in einem einfachen Deutsch, was jetzt genau aktuell die Situation ist um den Spieler und dessen Marktwert. (3. Interview, Pos. 22)

Eine andere Person wies darauf hin, dass der Stil sehr sachlich war, was dieser Person gefallen hat. Sie hat es folgendermassen formuliert:

Ich finde diese sachliche Zusammenfassung gut. Das gefällt mir. Es scheint ja wie Fakten basiert, diese Zusammenfassung, was mir gefällt. (2. Interview, Pos. 44)

Einzig eine interviewte Person äusserte sich dahingehend, dass sie den Stil etwas einfach fand und dieser nicht zum Lesen animierte. Sie formulierte es etwas salopp:

Es sieht aus, als hätte das eine Sekundar-Schülerin in einem Aufsatz geschrieben.
(lacht) (4. Interview, Pos. 54)

Formatierung

Die Meinungen zu der Formatierung waren ebenfalls einheitlich. Die meisten fanden diese gut, jedoch wurde darauf verwiesen, dass nicht mehr erwartet werden könnte an einen Fliesstext. Eine Person hat es wie folgt eingeschätzt:

Auch das Format finde ich eigentlich recht übersichtlich und ich glaube da kann man jetzt nicht viel mehr erwarten. (10. Interview, Pos. 47)

Die Gliederung in drei Absätze wurde dabei als sinnvoll bezeichnet. Eine Person beschrieb es folgendermassen:

Diese drei Absätze, in die es gegliedert ist sind auch sinnvoll, inhaltlich und auch leserfreundlich gestaltet. (8. Interview, Pos. 43)

Eine Person wünschte sich eine bessere Gliederung, anhand der Argumente. Sie hat es wie folgt beschrieben:

Vielleicht die Pro und Kontra ein bisschen besser gliedern, damit man das auch ja einfacher nachvollziehen kann. Dann wäre die ganze Sache sicher auch noch ein bisschen vertrauenswürdiger. (3. Interview, Pos. 34)

4.5.2 Änderungsbedarf

Aus den Interviews gingen der Bedarf für zwei Änderungen hervor. Es handelt sich dabei um Grafiken und Hervorhebungen und um eine Legende, welche die verwendeten Abkürzungen definiert.

Grafiken und Hervorhebungen

Zu möglichen Grafiken und Hervorhebungen liessen sich unterschiedlichste Meinungen ohne zentrale Tendenz feststellen. Einige Personen schätzten es als bereichernd ein, andere als überflüssig. Zum Teil empfanden die Personen das Einte als gewinnbringend aber wiederum das Andere nicht. Eine Person hat es folgendermassen eingeschätzt:

Besonders hervorheben ist nicht nötig, da ich ja nicht etwas Spezielles sehen will, das heraussticht, sondern ich will ja den ganzen Text lesen. Grafiken könnte schon spannend sein, muss nicht. (12. Interview, Pos. 48)

Dieselbe Person hat dazu folgendes ergänzt:

Wie sich die Kommentare der Beurteilung entwickeln, vielleicht sind jetzt die Kommentare eher beim Marktwert von so und so viel, ein halbes Jahr, ein Jahr später sind sie vielleicht doch dort, das wäre schon noch interessant. (12. Interview, Pos. 50)

Eine Person machte den Vorschlag, die Marktwertschätzungen noch optisch hervorzuheben. Sie hat es folgendermassen beschrieben:

Ich würde noch ergänzen, dass man die Marktwertschätzungen optisch hervorhebt. Ansonsten würde ich es so belassen. (7. Interview, Pos. 41)

Legende

Zwei Personen wiesen darauf hin, dass möglicherweise nicht alle Nutzerinnen und Nutzer die in der Zusammenfassung verwendeten Abkürzungen kennen. So wurde der Vorschlag angebracht, eine Legende zu diesen zur Verfügung zu stellen. Eine Person hat es wie folgt beschrieben:

Ja, vielleicht die Abkürzungen. Ich weiss nicht, ob alle diese Abkürzungen kennen, vielleicht gerade auch Nicht-Fussballkenner, dass man diese vielleicht ausschreibt oder nochmal erklärt. Sonst habe ich gar keine Verbesserungen. (11. Interview, Pos. 68)

4.6 Effizienz und Effektivität

4.6.1 Bewertung

Ziel der Nutzung

Das Ziel der Nutzung von Marktwerten generell hatte ein einheitliches Meinungsbild. Die Mehrheit der Interviewpersonen gaben an, dass sie die Marktwerte nutzen, um ein Bild oder eine Einschätzung über einen Spieler oder einen Verein bekommen. Eine Person hat es wie folgt beschrieben:

Die Marktwerte nutze ich zum der Entwicklung nachzugehen oder verschiedenen Werte zu sehen, wie viel hat dieser Spieler wert, um die Unterschiede zu erkennen. (1. Interview, Pos. 5)

Andere Personen sagten aus, dass sie die Marktwerte nutzten, um Vergleiche zu ziehen. Eine Person hat es folgendermassen beschrieben:

Mehrheitlich eigentlich, um Vergleiche zu ziehen. (I: Also zwischen den Spielern?) Genau, also zwischen Spielern und Mannschaften. (I: Okay) Allenfalls auch Ligen. (3. Interview, Pos. 6)

Nachvollziehbarkeit ohne Zusammenfassung

Die grosse Mehrheit von 9 Personen hat angegeben, dass derzeitige die Entstehung Marktwerte schwer nachvollziehbar oder mit Aufwand verbunden ist. Eine Person hat es wie folgt geschildert:

Für mich persönlich war es nie wirklich nachvollziehbar. Ich habe mich auch keine Gedanken gemacht, wie die Marktwerte zustande kommen. Das Einzige, was ich weiss, ist, dass es diese Diskussion gibt und anhand von diesen ungefähr der Marktwert bestimmt wird. (11. Interview, Pos. 10)

Zwei Personen haben angegeben, dass es zwar nachvollziehbar ist aber die Gewinnung der nötigen Informationen viel Aufwand benötigt. Eine Person hat es folgend beschrieben:

Wenn man die Kommentare durchliest, dann ist es nachvollziehbar. Man kann das ganze Geschehen verfolgen und die Pros und Kontras abwägen. Und ja anhand der Leistungsdaten ist es sehr nachvollziehbar. (7. Interview, Pos. 10)

Antizipierte Nutzung

Die grosse Mehrheit von elf Interviewpersonen hat angegeben, dass sie eine Zusammenfassung tatsächlich nutzen würden. Die meisten fanden die Zusammenfassung gewinnbringend, da sie den Aufwand gering hält und viele Informationen liefert. Eine Person hat es wie folgt geschildert:

Ich denke, es würde mich nichts daran hindern, einfach die Zusammenfassung noch anzuschauen. Eventuell, wenn ich jetzt nur kurz Marktwerte vergleichen möchte und mir schnell ein Bild machen würde, ohne das Ganze zu hinterfragen. (6. Interview, Pos. 59)

Eine Person hat Vorbehalte geäussert, könnte sich die Nutzung aber dennoch vorstellen. Sie hat es folgend erläutert:

Ja, wenn ich nicht daran interessiert bin, zu wissen, wie dieser Marktwert entstanden ist oder wenn ich von vorn rein absolut derselben Meinung wäre, würde ich wahrscheinlich nicht die Marktwertdiskussion lesen, weil ich keinen Grund hätte, daran zu zweifeln. Vor allem der Zweifel treibt mich eigentlich dazu, die Marktwertdiskussion zu analysieren, damit ich verstehen kann, wie es zum Wert

kommt. Wenn ich keine Zweifel daran hätte, würde ich es nicht benutzen, dieses Feature. (8. Interview, Pos. 63)

4.7 Weitere Anwendungen

Gerüchteküche

Bei den weiteren Anwendungen wurden zwei konkrete Vorschläge gemacht. Der einte bezieht sich auf die Zusammenfassungen vom Forum der Gerüchteküche. Eine Person hat es folgend beschrieben:

Ja, das kann ich mir gut vorstellen. Vielleicht vor allem in Bezug auf die Transfermarktgerüchte, also die Gerüchteküche. Dort wird sicher sehr eine gute Anwendung finden, solche Zusammenfassungen. (1. Interview, Pos. 82)

Statistiken und Leistungsdaten

Der andere Vorschlag bezog sich auf die Zusammenfassungen von diversen Statistiken oder Leistungsdaten. Eine Interviewperson hat es folgendermassen beschrieben:

Ich denke, um Vergleiche zu machen. Also die Fussballwelt besteht aus Vergleichen von Spielern und ich denke, es wäre toll, einen Vergleich, z.B. Messi im Alter von 17 und Jamal im Alter von 17, die Leistungsdaten mit KI so zu vergleichen, wäre sehr spannend. (4. Interview, Pos. 86)

5 Diskussion

In diesem Kapitel werden die zentralsten Ergebnisse der Forschung zusammengefasst und mit Einbezug der Theoretischen Grundlagen diskutiert. Anschliessend wird auf die Limitation dieser Arbeit eingegangen, ein Fazit und ein Ausblick gegeben.

5.1 Zusammenfassung und Interpretation der Forschungsergebnisse

Im Ergebnisteil konnten zu den sechs relevanten Aspekten differenzierte Beurteilungen und Änderungsbedarf erläutert werden. Mit dem letzten Abschnitt der Ergebnisse konnte ebenfalls ein Ausblick für potenzielle weitere Anwendungsbereiche gegeben werden. Auch wenn die Erkenntnisse nicht direkt relevant sind für die Beantwortung der Fragestellung, sind sie innerhalb des menschenzentrierten Gestaltungsprozesses interessante Erkenntnisse, welche beim Start des Entwicklungsprozesses von möglichen weiteren ähnlichen Anwendungen berücksichtigt werden können.

Beim Aspekt Genauigkeit konnte eine gewisse Zufriedenheit festgestellt werden. Die Mehrheit war mit den enthaltenen Informationen zufrieden und fanden auch die Reichhaltigkeit passend. Einzelne Personen waren jedoch weniger zufrieden damit. Auch beim Änderungsbedarf konnte nur eine Subkategorie gebildet werden. Diese Ergebnisse decken sich mit denjenigen von Arnold & Hörauf (2024), die eine hohe Kohärenz und semantische Übereinstimmung bei KI-generierten Zusammenfassungen feststellen konnten.

Beim Umfang der Zusammenfassung konnte eine einheitliche positive Beurteilung festgestellt werden. Konträr die dieser Beurteilung machten gesamthaft vier der interviewten Personen die Änderungsvorschläge, zwei unterschiedlich lange Versionen zur Verfügung zu stellen oder die Länge der Zusammenfassung anhand der Länge des Forums zu bestimmen. Den Umfang der Zusammenfassung anhand des Inputmaterials zu bestimmen, deckt sich mit der Erkenntnis von Ter Hoeve et al. (2022), die auf die Relevanz der Berücksichtigung des Inputmaterials bei KI-generierten Zusammenfassungen hinwiesen. So sollte nicht nur der Umfang des Inputmaterials berücksichtigt werden, sondern auch die Komplexität dessen (Ter Hoeve et al., 2022).

Bei den Ergebnissen zum Vertrauen hinterfragen einige die Qualität der Input-Daten. Im Rahmen einer Zusammenfassung ist dies schwer zu optimieren. Die Relevanz der Input-Daten deckt sich mit den Erkenntnissen von Ter Hoeve et al. (2022). Eine beträchtliche Anzahl an Personen gaben an, dass sie Verweise auf das Diskussionsforum bereichernd finden. Dies ist in der Tat eine interessante Anpassung, die auch mehr Personen zum Forum bringen könnte. Auch die Vorschläge zur Einbindung von statistischen Werten sind interessant, da Transfermarkt bereits eine Fülle dieser Werte bieten.

Zu Transparenz und Kontrolle konnte festgestellt werden, dass ein Bedarf danach besteht, über die Entstehung der Zusammenfassung im Bilde zu sein. Dennoch war das persönliche Interesse danach teilweise gering. Ähnlich ist es beim Bedarf nach weiteren Kontrollmöglichkeiten, dieser fiel ebenfalls eher gering aus. Die verlangte Effizienz und die Rolle als Rezipient scheinen im Vordergrund zu stehen. Diese Erkenntnis deckt sich mit derjenigen von Cheng et al. (2022), die ebenfalls einen negativen Zusammenhang zwischen Effizienz und Transparenz und Kontrolle feststellten.

Bei Stil und Format konnte festgestellt werden, dass bei der Zusammenfassung ein angemessener Sprachstil verwendet wurde. Im Gegensatz zu Cheng et al. (2022) konnten keine grossartig differenzierten Bedürfnisse gegenüber dem sprachlichen Stil gefunden werden. Bei möglichen Grafiken oder Hervorhebungen wurde jedoch ein breites Meinungsbild vorgefunden. Dies deckt sich mit den Erkenntnissen von Ter Hoeve et al. (2022). Ansonsten konnte die eher schlichte Formatierung überzeugen. Auch wenn angemerkt wurde, dass es bei einem Fliesstext wenig Optimierungsspielraum gibt.

Bei der Effektivität und Effizienz zeigte sich ein sehr einheitliches Bild. Alle interviewten Personen würden eine Zusammenfassung von Marktwertdiskussionen tatsächlich nutzen. Eine Person davon äusserte noch einige Vorbehalte. Die Mehrheit empfand die Zusammenfassung als Aufwandssparend und gewinnbringend. Die Ergebnisse sind erfreulich, es sollte jedoch bedacht werden, dass die Interviewpersonen diverse Änderungsvorschläge gemacht haben und auf Unzulänglichkeiten hinwiesen. Diese sollten dennoch nicht vernachlässigt werden.

Bei den weiteren Anwendungen gab es die Zusammenfassung der Gerüchteküche und von Statistiken, die vorgeschlagen wurden. Insbesondere die Zusammenfassung der Gerüchteküche zu Transfers ist vom Aufbau ähnlich wie die Foren zu den Marktwerten. Daher ist davon auszugehen, dass die Ergebnisse weitgehend übertragen werden können.

5.2 Kritische Würdigung und Fazit

Kritisch zu betrachten ist die Tatsache, dass sich einige der untersuchten relevanten Aspekte gegenseitig beeinflussen und nicht genau voneinander abzutrennen sind. So beeinflusst beispielsweise die Wahrnehmung der Transparenz die des Vertrauens und umgekehrt (Cheng et al., 2022). Ähnliches konnte bei den Ergebnissen dieser Arbeit festgestellt werden, beispielsweise, dass eine als passend wahrgenommene Zusammenfassung das Bedürfnis nach Kontrolle minderte. Offensichtlich ist auch der Zusammenhang zwischen Genauigkeit und Umfang. Durch die Aufteilung dieser Aspekte konnte dennoch vertiefte Erkenntnisse über diese gewonnen werden.

Ein weiterer kritisch zu betrachtender Punkt ist der Datenschutz. Wie aus der Arbeit hervorging, vertrauen nicht alle Nutzerinnen und Nutzer im selben Ausmass in KI. Diese Erkenntnis deutet darauf hin, dass nicht alle Nutzenden damit einverstanden sein könnten, wenn ihre Forumsbeiträge mit KI weiterverarbeitet werden.

Ein weiteres Problem könnte die rasante Entwicklung und stetige Veränderung von KI darstellen. So hatte beispielsweise OpenAI im Verlauf dieser Arbeit ein neues Modell und ein weiteres Research-Preview veröffentlicht. KI kann mit demselben Prompt aber leicht unterschiedlichem Input-Material unterschiedliche Ergebnisse liefern. Durch die stetige Weiterentwicklung kann dies dazu führen, dass ein Prompt ebenfalls stetig angepasst werden muss. Zudem verbreitet sich die Anwendung von KI stetig weiter, was die Ergebnisse dieser Arbeit beispielsweise zum Vertrauen verändern könnte.

Dennoch konnten in dieser Arbeit relevante Forschungsergebnisse zu zentralen Aspekten bei KI-generierten Zusammenfassungen von Forumdiskussionen erarbeitet werden. Die Erkenntnisse basieren auf der Beurteilung von Nutzerinnen und Nutzern von Transfermarkt.de und wurden ausgiebig gegenüber den Erkenntnissen aus dem Theorieteil diskutiert. Die Ergebnisse bilden eine Grundlage für den weiteren Entwicklungsprozess und sie bieten einen substanziellen Erkenntnisgewinn für das Entwicklungsteam von Transfermarkt. Zudem leisten sie einen Beitrag für erfolgreiche KI-Anwendungen, welche anhand der Nutzerinnen und Nutzer und nicht an ihnen vorbei entwickelt werden.

5.3 Ausblick

Im Nachgang dieser Arbeit wäre es sinnvoll, die Erkenntnisse im Sinne des menschenzentrierten Gestaltungsprozesses in die weitere Entwicklung einfließen zu lassen und eine erneute Iterationsschleife zu starten. In diesem Sinne könnte ein erstes Mockup erstellt werden, welches unter Einbindung der Nutzerinnen und Nutzer weiter evaluiert werden kann.

Ein Teil der Erkenntnisse kann zudem auf anderweitige, nicht KI-generierte Zusammenfassungen übertragen werden. Die Beurteilung zur Genauigkeit und zum Umfang ist beispielsweise nicht geprägt von der Tatsache, wer oder was diese Zusammenfassung erstellt hat. Somit können diese Erkenntnisse teilweise auf menschliche oder algorithmische Zusammenfassungen übertragen werden.

Ähnlich sind die Erkenntnisse zum Vertrauen und der Transparenz einzustufen. Diese sind sehr geprägt durch die Tatsache, dass KI zum Einsatz gekommen war. Somit können die Ergebnisse einen spannenden Einblick in die Vertrauensbildung und die Anforderungen an Transparenz bei KI-generierten Inhalten bieten.

6 Literaturverzeichnis

- 20 Jahre Marktwerte: Vom „reinen Hobby-Projekt“ zur relevanten Größe im Weltfußball. (2024, 10. Dezember). *Transfermarkt*. Zugriff am 3.6.2025. Verfügbar unter:
<https://www.transfermarkt.ch/20-jahre-marktwerte-vom-reinen-hobby-projekt-quot-zur-relevanten-grosse-im-weltfussball/view/news/445024>
- Arnold, J. B. & Hörauf, D. (2024). A Multimetric Approach for Evaluation of ChatGPT-Generated Text Summaries. *IEEE Engineering Management Review*, 52(3), 43–53.
<https://doi.org/10.1109/emr.2024.3381176>
- Bracamonte, V., Hidano, S., Tesfay, W. B. & Kiyomoto, S. (2020). User Study of the Effectiveness of a Privacy Policy Summarization Tool. *Communications in computer and information science* (S. 186–206). https://doi.org/10.1007/978-3-030-49443-8_9
- Cheng, R., Smith-Renner, A., Zhang, K., Tetreault, J. R. & Jaimes, A. (2022). Mapping the Design Space of Human-AI Interaction in Text Summarization. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.2206.14863>
- Datenpflege-"How-to-Guide". (n. d.). *Transfermarkt*. Zugriff am 3.6.2025. Verfügbar unter:
<https://www.transfermarkt.ch/intern/faq>
- Deutsches Institut für Normung e.V. EN ISO 9241-210 (DIN EN ISO 9241-210). (2011). *Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*. Berlin: Beuth.
- deutschlandfunk.de. (2020, 20. August). Der Preis des Erfolgs. *Deutschlandfunk*. Zugriff am 3.6.2025. Verfügbar unter: <https://www.deutschlandfunk.de/der-preis-des-erfolgs-100.html>
- Flick, U. (2019). *Qualitative Sozialforschung - Eine Einführung* (9. Auflage.). Reinbek bei Hamburg: Rowohlt.
- Ganesan, K. (2018). ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1803.01937>
- Helfferrich, C. (2011): *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. 4. Auflage, Wiesbaden: VS Verlag für Sozialwissenschaften. Rädiker und Kuckartz (2019)

- Kambhamettu, H., Flores, J. & Head, A. (2024). Traceable Text: Deepening Reading of AI-Generated Summaries with Phrase-Level Provenance Links. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.2409.13099>
- Kreutzer, R. T. (2023). *Künstliche Intelligenz verstehen*. Springer eBooks.
<https://doi.org/10.1007/978-3-658-42598-2>
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (Grundlagentexte Methoden, 4., überarbeitete Auflage). Weinheim: Beltz.
- Kuckartz, U. & Rädiker, S. (2024). *Fokussierte Interviewanalyse mit MAXQDA*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-40212-9>
- Marktwert-Definition - So funktioniert's - Marktwertanalyse - Forum | Seite 1. (2021, 13. Oktober). *Transfermarkt*. Zugriff am 3.6.2025. Verfügbar unter: https://www.transfermarkt.ch/marktwert-definition-so-funktioniert/thread/forum/127/thread_id/26444
- Moser, C. (2012). *User Experience Design*. Berlin Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-13363-3>
- Sarodnick, F. & Brau, H. (2016). *Methoden der Usability Evaluation* (3. Aufl.). Bern: Hogrefe.
- Silvagnoli, L. M., Shepherd, C., Pritchett, J. & Gardner, J. (2021). Optimizing Readability and Format of Plain Language Summaries for Medical Research Articles: Cross-sectional Survey Study. *Journal Of Medical Internet Research*, 24(1), e22122. <https://doi.org/10.2196/22122>
- Soni, M. & Wade, V. (2023). Comparing Abstractive Summaries Generated by ChatGPT to Real Summaries Through Blinded Reviewers and Text Classification Algorithms. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.17650>
- Stiglic, G., Kopitar, L., Gosak, L., Kocbek, P., He, Z., Chakraborty, P. et al. (2023). Improving Primary Healthcare Workflow Using Extreme Summarization of Scientific Literature Based on Generative AI. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2307.15715>
- Ter Hoeve, M., Kiseleva, J. & Rijke, M. (2022). What Makes a Good and Useful Summary? Incorporating Users in Automatic Summarization Research. *Proceedings Of The 2022 Conference Of The North American Chapter Of The Association For Computational*

Linguistics: Human Language Technologies, 46–75. <https://doi.org/10.18653/v1/2022.naacl-main.4>

What is Human-Centered AI (HCAI)? (2025, 13. Mai). *The Interaction Design Foundation*. Verfügbar unter: <https://www.interaction-design.org/literature/topics/human-centered-ai?srsId=AfmBOoruFFMbeRVGpo2MkGVFETrMFuweHZdFhbpEjRnTTuOJPtmsy7Od>

Yamamoto, S., Suzuki, R., Fukusato, T., Kataoka, H. & Morishima, S. (2021). A Case Study on User Evaluation of Scientific Publication Summarization by Japanese Students. *Applied Sciences*, 11(14), 6287. <https://doi.org/10.3390/app11146287>

7 Abbildungsverzeichnis

Abbildung 1. Menschenzentrierter Gestaltungsprozess, eigene Darstellung, nach DIN EN ISO 9241-210 (2011)	4
Abbildung 2. Untersuchungsdesign (eigene Darstellung)	9
Abbildung 3. Kategoriensystem (eigene Darstellung)	14

8 Tabellenverzeichnis

Tabelle 1. Ausschnitt aus Interviewleitfaden	11
Tabelle 2. Transkriptionsregeln (in Anlehnung an Rädiker und Kuckartz, 2019, S. 44-45)	13

9 Hilfsmittelverzeichnis

KI- Assistenzsystem	Teile / Stellen in der Arbeit	Einsatz
ChatGPT	Theorie	Hilfe bei möglichen Keywords für die Literaturrecherche
ChatGPT	Methoden / 3.2.3	Erstellung der prototypischen Zusammenfassung
Turboscribe	Datenauswertung	Vortranskription der Audiodateien der Interviews

10 Anhangsverzeichnis

A: Ausschnitt Marktwert Diskussionsforum

B: Prototypische Zusammenfassung

C: Interviewleitfaden

D: Definition Kategorien

E: Transkripte