

Erfolgsfaktoren von Augmented-Reality-Applikationen: Analyse von Nutzerrezensionen mit dem Review-Mining- Verfahren

Laura Sophie Gravemeier¹, Lisa Berkemeier¹, Oliver Thomas¹

¹ Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Osnabrück,
Deutschland

{lgravemeier,lisa.berkemeier,oliver.thomas}@uni-osnabrueck.de

Abstract. Nachdem die Leistungsentwicklung von mobilen Endgeräten der letzten Jahre die Nutzung von Augmented-Reality-Anwendungen dem Massenmarkt zugänglich gemacht hat, sind diese nun gängiger Bestandteil der App Stores. Es besteht jedoch eine Forschungslücke im Hinblick auf klare Gestaltungsrichtlinien für Augmented-Reality-Applikationen. Rezensionen in den App Stores bieten an dieser Stelle eine wertvolle Einsicht in das Nutzererlebnis. Zur Identifikation von Erfolgsfaktoren wird ein Review Mining mit den Nutzerkommentaren zu 20 Augmented Reality Applikationen durchgeführt. Durch eine Sentiment Analysis werden positive und negative Nutzererfahrungen voneinander getrennt, um darin wiederholt auftretende Lob- und Kritikpunkte durch ein Topic Modeling zu erkennen. Die Erkenntnisse erweitern die Wissensbasis durch die Identifikation von Adoptionstreibern und Entwicklungsfeldern für die Gestaltung von Augmented-Reality-Applikationen. Insbesondere technische Zuverlässigkeit und Kompatibilitätsprobleme wurden als Entwicklungsbedarf identifiziert, während die volle Ausnutzung des Potenzials des Augmented Reality Interface Designs sich als klarer Treiber zeigt.

Keywords: Augmented Reality, User Experience, Review Mining, Sentiment Analysis, Topic Modeling

1 Einleitung

Applikationen (Apps) können das Konzept Augmented Reality (AR) auf verschiedenen Endgeräten umsetzen, indem sie digitale Objekte in die reale, wahrgenommene Umgebung des Nutzers integrieren. Nach der erfolgreichen Einführung der App „Pokémon Go“ werden nun verschiedene AR-Apps entwickelt und konsumiert. Neben der Anwendung in der Spieleindustrie werden AR-Apps auch in klassischen Branchen wie der Möbelindustrie und im Bildungssektor entwickelt. Hierzu stellte Apple Inc. [1] auf der World Wide Developer Conference im Juni 2017 das ARKit vor, im September antwortete Google LLC [2] im Rahmen des Google Developers Day mit der Ankündigung des ARCore. Sie dienen als Entwicklungswerkzeuge für AR-Apps und bereiten den Weg für die weitere Verbreitung von Mobile-Augmented-Reality-Apps (MAR-Apps) in den gängigen App Stores.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Obwohl die ersten Überlegungen zu Augmented bzw. Virtual Reality bereits lange zurückliegen [3], weitet sich damit erst jetzt ihre Verwendung auf eine Vielzahl von Endnutzern aus. Folglich befindet sich die wissenschaftliche Forschung zu Design und Konzeption von MAR-Apps mit Berücksichtigung des Nutzerverhaltens und deren Umsetzung derzeit noch in ihren Anfängen [4–6]. Allerdings ist die Erhebung von Nutzerdaten durch Usability- und User-Experience-Experimente mit hohen Kosten verbunden und bisher vorliegende Untersuchungen weisen häufig noch Qualitätsdefizite auf [7]. Für herkömmliche Apps stehen zwar bereits Richtlinien zur Interfacegestaltung auf Basis wissenschaftlicher Forschung zur Verfügung (z.B. [8]), allerdings können diese aufgrund der besonderen Interaktionsmöglichkeiten nur begrenzt auf AR-Apps übertragen werden.

Die Technologie eröffnet neue Möglichkeiten zur Dienstleistungserbringung seitens der Industrie und schafft neue Nutzungserlebnisse für den Endkunden. Diese Erfahrungen werden in Form von Bewertungen auf entsprechenden Plattformen, wie den App-Stores von Apple und Google, rezensiert. Diese unstrukturierten Daten können wertvolle Erkenntnisse für die Entwicklung von AR-Apps beinhalten. Ziel dieses Beitrags ist die Identifikation von Erfolgsfaktoren, aufgegliedert in Adoptionstreiber bzw. Entwicklungsfelder, die aus Rezensionen mittels Text Mining abgeleitet und analysiert werden. Die Untersuchung wird strukturiert durch die Forschungsfrage (FF) *Welche Aspekte von AR-Apps werden wiederholt positiv bzw. negativ zurückgemeldet und können daher als Erfolgsfaktoren von AR-Apps gewertet werden?* Dazu werden in Kapitel zwei zunächst die theoretischen Grundlagen der AR und des Review Mining (RM) als zentrale Konzepte erläutert. Anschließend wird in Kapitel drei die Durchführung und verwendete Methoden des RM beschrieben, um eine Rekonstruierbarkeit zu gewährleisten. In Kapitel vier werden die erzielten Ergebnisse umfassend dargestellt und in Kapitel fünf diskutiert. Abschließend wird in Kapitel sechs ein Fazit gezogen und ein Ausblick auf weiteren Forschungsbedarf gegeben.

2 Theoretische Grundlagen

2.1 Anwendungsgebiete und Interfaces von Augmented Reality Applikationen

AR beschreibt die Integration virtueller Elemente in die reale Umwelt. Der nahtlose Übergang virtueller Objekte in die reale Umgebung wird als Erweiterung der Realität verstanden. Dabei wird eine vermischte 3D Realität erzeugt, mit der in Echtzeit interagiert werden kann [9, 11]. MAR, d.h. die Verwendung von AR auf mobilen Geräten [12], wird von Smartphones geprägt [5], [13]. Smartphones stehen einer breiten Masse an Anwendern zur Verfügung und können im Gegensatz zu spezifischen Endgeräten, wie AR-Brillen, die Weitergabe der AR-Technologie von einer forschungszentrierten Entwicklung hin zum Endnutzermarkt beschleunigen. AR-Apps werden in der Regel nach Anwendungsfeldern kategorisiert [vgl. 4]. Über verschiedene Klassifikationen hinweg finden sich AR-Apps mit Bezug zu Bildung, Gesundheit, Marketing, Bauwesen und Fertigung, Gaming und Entertainment, Militär, Visualisierung, sowie Navigation und Tourismus [4], [9], [14–17].

Im Rahmen dieser Anwendungsgebiete werden AR-Apps sich nur durchsetzen können, wenn sie im Vergleich mit herkömmlichen Apps Vorteile bieten. Dabei nutzt AR das Potential der Immersion, um das Nutzererlebnis zu verbessern. So wird z.B. beim Online-Shopping mit der IKEA AR-App die Anordnung virtueller Produkte im tatsächlichen Raum und somit eine bessere Entscheidungssituation ermöglicht. Die vielfältigen Interaktionsmöglichkeiten der AR-Technologie ermöglichen neue User Interfaces (UI). Insbesondere tragbare MAR-Geräte können als „magic lens“ [17] bezeichnet werden, durch die der Benutzer die Realität erweitert wahrnimmt. Die Interaktion mit den darin sichtbaren virtuellen Objekten kann in einem tangible UI [18] direkt durch die Manipulation der korrespondierenden realen Objekte erfolgen. Zusätzlich dient ein kollaboratives UI mehreren Benutzern als Basis für einen ausgeführten Prozess, sodass gemeinsam nicht nur mit realen, sondern auch mit virtuellen Objekten interagiert werden kann [19], [20]. Generell wird die Zusammenführung verschiedener Interaktionsmöglichkeiten und somit ein hybrides UI angestrebt, um eine möglichst realitätsnahe Erfahrung zu erzeugen [17]. Um die Schnittstelle zwischen Anwender und AR-System noch lebendiger zu gestalten, kommen multimodale UIs zum Einsatz. Hier werden mehrere Sinne des Benutzers für die Verarbeitung des In- und Outputs des AR-Systems verwendet [17].

2.2 Review Mining

Die Analyse von Reviews zur Zusammenfassung verschiedener Produkteigenschaften und Nutzereindrücke wird abgeleitet vom Text Mining als RM bezeichnet [21] und als Tool zur Analyse von Apps eingesetzt [22], [23]. Durch die Betrachtung von Einschätzungen der Nutzer stehen vor allem die Methoden des Opinion Mining bzw. der Sentiment Analysis (SA) im Vordergrund. Die Abgrenzung der Begriffe Opinion Mining und SA ist umstritten [24], allerdings dienen beide der Analyse von Texten mit subjektiven Inhalten im Gegensatz zu klassischen Text Mining Anwendungen [21], [25].

Eine zentrale Komponente der Analyse subjektiver Daten ist die Einschätzung ihrer Polarität bzw. der Richtung, der darin ausgedrückten Empfindungen. Diese Klassifikation kann durch verschiedene Methoden erreicht werden: mithilfe von Lexika, in denen Sentiment-Werte für enthaltene Wörter bereits definiert sind, aber auch gestützt durch maschinelle Lernverfahren wie das Natural Language Processing [25]. Diese Lernverfahren benötigen bereits klassifizierte Datensätze, um davon eine systematische Einteilung abzuleiten. Um möglichst genaue Ergebnisse zu erzielen, werden domänenspezifische Trainingsdaten benötigt. Allerdings ist die manuelle Generierung individueller Datensätze für jedes neue Anwendungsgebiet mit einem hohen Aufwand verbunden und daher werden auch domänenübergreifende Ansätze verfolgt [26]. Meist wird die Polaritätseinschätzung auf Basis des gesamten Textdokuments durchgeführt, es kann jedoch auch auf Ebene von Sätzen oder sogar Aspekten differenziert werden. Die erhöhte Granularität ermöglicht auf der einen Seite eine klarere Zuordnung der subjektiven Einschätzung zum Meinungsgegenstand, kann aber durch den Verlust von Kontext auch unerwünschte Effekte hervorrufen [27].

Zusätzlich zur Polarität der ausgedrückten Meinungen, muss auch der inhaltliche Gegenstand der Reviews ausgewertet werden. Um wiederkehrende Themen und ihre Zusammenhänge zu kategorisieren, werden Topic Modeling Methoden genutzt [28], [29]. Häufig wird Latent Dirichlet Allocation (LDA) angewandt, das übergeordnete Themen und assoziierte Inhalte in Textquellen identifiziert [30], [31]. Klassische LDA Verfahren fordern dabei eine Angabe der Topicanzahl, während das hierarchische LDA diese dynamisch selbst erlernt [32].

3 Analyse von Rezensionen über Augmented Reality Applikationen mit dem Review-Mining-Verfahren

Im Rahmen dieser Arbeit wird ein RM mithilfe der oben erläuterten Methoden durchgeführt. Hier wird zunächst der konkrete Untersuchungsablauf beschrieben, der zum Überblick in Abbildung 1 dargestellt ist. Zuerst wurden die zu untersuchenden Daten ausgewählt und extrahiert. Mit dem Ziel möglichst viele Reviews zu erfassen, wurden Daten aus den Apple App Stores aller Nationen mit Englisch als Muttersprache berücksichtigt: die Vereinigten Staaten, Kanada, das Vereinigte Königreich, Irland, Australien, Neuseeland und Südafrika [33]. Als Indikator für eine möglichst hohe Reviewdichte wurden die 20 Apps mit der höchsten Anzahl an Benutzerbewertungen im App Store der Vereinigten Staaten ausgewählt. Daten zu Download- oder Reviewanzahl können im App Store nicht eingesehen werden. Die Extraktion der Daten wurde mittels eines PHP-Skripts (modifiziert nach [34]) durchgeführt, das Datum, Rating, Titel, Inhalt und weitere Metadaten des Reviews gescreapt hat.

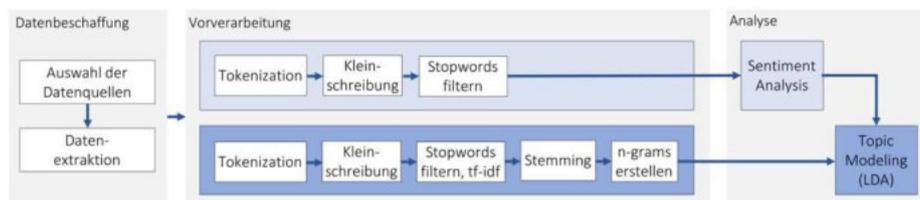


Abbildung 1. Ablauf des durchgeführten RM

Zur Analyse der Daten wurde das Tool RapidMiner Studio (Version 8.2.000) verwendet. Im ersten Schritt wurden durch das Scraping oder Spam entstandene Duplikate entfernt, wonach mit 130.262 Datensätzen gearbeitet werden konnte. Deskriptive Daten zur initialen Stichprobe sind in Tabelle 2 im Anhang dargestellt.

Als Grundlage der SA dienen die textuellen Bestandteile der gescreapten Daten, d.h. Reviewtitel und -inhalt. Diese wurden in ihre einzelnen Wortbestandteile aufgeteilt (Tokenization) und in Kleinbuchstaben umgewandelt. Zusätzlich wurden sogenannte Stopwords herausgefiltert, d.h. in der englischen Sprache grundsätzlich gehäufte Worte oder auf den App-Inhalt bezogene Worte ohne Relevanz für die ausgedrückte Meinung. Anschließend wurde auf Basis der verbliebenen Tokens eine SA durchgeführt. Da kein

domänenspezifisches Training Set zur Verfügung stand, wurde ein Sentiment Score mit dem domänenübergreifenden Tool SentiWordNet 3.0 [35] berechnet.

Vorbereitend für das Topic Modeling Verfahren wurde der Datensatz zunächst verkleinert, um eine ausgewogene Zusammensetzung der Daten zu erreichen. Zuvor lag der Reviewanteil der drei führenden Apps bei 91,18% und Pokémon Go dominierte mit 86.217 von 130.262 Dateneinträgen. Um einen zu starken Einfluss dieser drei Apps zu verhindern und gleichzeitig die Stichprobe nicht unnötig zu verkleinern, wurden per Zufallsauswahl jeweils nur 5.000 ihrer Reviews zur Untersuchung ausgewählt und damit insgesamt noch 26.495 Reviews berücksichtigt. Anhand des Sentiment Scores wurden drei Gruppen gebildet: positive (Sentiment Score ab 0,2), negative (Sentiment Score bis -0,2) und neutrale Reviews. Der Sentiment Score nimmt keine kategorische Einteilung vor und somit wurden für eine möglichst fehlerfreie Einteilung die positive und negative Gruppe nicht im Nullpunkt der Sentimentskala getrennt. Die Vorverarbeitung der Daten umfasste ebenfalls den Prozess der Tokenization, die Umwandlung in Kleinbuchstaben und das Herausfiltern von Stopwords. Zusätzlich wurde das „Term Frequency-Inverse Document Frequency“-Verfahren angewandt, um nicht untersuchungsrelevante Terme zu entfernen. Daraufhin wurden die Tokens auf ihre Stammform reduziert und in N-Gramme mit einer maximalen Länge von 3 zerlegt. Anschließend wurde ein Topic Modeling per LDA angewandt. Da kein hierarchisches LDA Tool zur Verfügung stand, wurden Verteilungen mit 5 bis 30 Themen getestet, um interpretierbare und möglichst aussagekräftige Topic Models zu erhalten. Dabei wurden in 1.000 Iterationen reproduzierbare Topic Models mit jeweils 15 zugehörigen assoziierten Termen (Top Words) pro Topic erzeugt. Für die Bestimmung der Hyperparameter α bzw. β wurde eine Heuristik angewandt und sie wurden automatisch iterativ optimiert, was der gängigen Praxis entspricht [28].

Um möglichen Qualitätsdefiziten des Datensatzes entgegenzuwirken, wurden weitere Ausschlusskriterien durchgesetzt. Als Ausgleich möglicher Ungenauigkeiten in der SA durch das domänenübergreifende Vorgehen wurden die Ratings als weiterer Indikator für die Einteilung der finalen Untersuchungsgruppen genutzt. Aus den positiven Reviews wurden danach nur noch Reviews mit 4 oder mehr Sternen, aus den negativen Reviews Einträge mit 2 oder weniger Sternen betrachtet.

4 Identifikation von Erfolgsfaktoren für AR-Applikationen

4.1 SA und Teilung des Datensatzes in Untersuchungsgruppen

Nach dem beschriebenen Verfahren wurde die Stichprobe anhand einer SA in drei Gruppen geteilt. Als Indikator für die Validität der durchgeführten SA wird zunächst der Zusammenhang zwischen Sentiment Score und dem vergebenen Rating des Reviews betrachtet. Das Rating wird auf einer Skala von 1 (schlechteste Bewertung) bis 5 (beste Bewertung) vergeben. Ein hoher Sentiment Score geht mit einem hohen Rating einher, $r(26.495) = 0,377$, $p < 0,001$ und Tabelle 1 zeigt klare deskriptive Unterschiede im durchschnittlichen Rating zwischen den Untersuchungsgruppen.

Tabelle 1. Rating und Sentiment Score der vorläufigen Untersuchungsgruppen

	<i>Reviews in Stichprobe</i>	<i>Negative Gruppe</i>	<i>Neutrale Gruppe</i>	<i>Positive Gruppe</i>
Anzahl	26.495	1.330	15.291	9.874
Durchschnittliches Rating	4,079	3,009	3,777	4,690
Durchschnittliches Sentiment	0,143	-0,357	0,05	0,354

Insgesamt weist dies auf eine erfolgreiche Einteilung des Sentiments hin. Allerdings enthält die negative Untersuchungsgruppe viele Reviews mit positivem Rating (siehe Abbildung 2). Da von einem positiven Inhalt eines Reviews mit 5 Sternen auszugehen ist, könnten diese einen störenden Einfluss auf die Ergebnisse nehmen. Daher wurde die positive bzw. negative Untersuchungsgruppe um Reviews mit zu niedrigen bzw. hohen Ratings gekürzt. Nach dieser Einteilung werden die Untersuchungsgruppen separat betrachtet und davon ausgegangen, dass ihr Inhalt Aufschluss über Erfolgsfaktoren von AR-Apps geben kann.

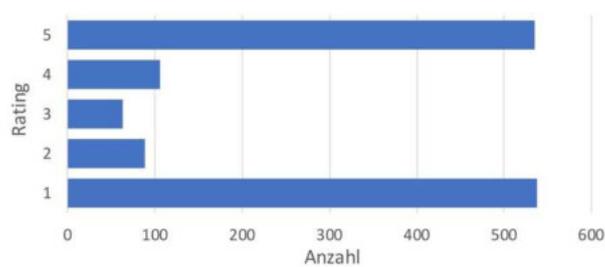


Abbildung 2. Rating der Reviews in negativer Untersuchungsgruppe

4.2 Topic Modeling in negativer Untersuchungsgruppe

Die Untersuchungsgruppe mit negativen Reviews umfasst nach dem oben beschriebenen Ausschlussverfahren 626 Datensätze, davon gehören 397 zur App „Pokémon Go“. Mit einem durchschnittlichen Rating von 2,669 und Sentiment Score von 0,014 im ungekürzten Datensatz gehört diese App nach beiden Kriterien zu den schlechtesten. Eine starke Repräsentation im negativen Datensatz ist damit nicht zu vermeiden, jedoch wird bei der Interpretation eine Abstrahierung von App-Inhalten angestrebt. Als Grundlage für die Auswertung dient ein per LDA generiertes Topic Model mit 10 vorgegebenen Topics (siehe Tabelle 3 im Anhang), wobei sinnvolle Top Words in Kombination mit hochassozierten Reviews interpretiert werden.

Die Betrachtung dieses Topic Modells fasst wiederkehrende Inhalte in negativen Reviews zusammen und macht somit Problemfelder deutlich. Hierunter befinden sich auch einige Themen, die keinen eindeutigen Bezug zu AR haben. So wird allgemeine Unzufriedenheit ausgedrückt mit beispielhaften Reviewinhalten wie „huge regret downloading“. Zudem wird das Preis-Leistungs-Verhältnis in negativen Reviews als

nicht ausgeglichen wahrgenommen und die jeweilige App z.B. als „a waste of money“ bezeichnet. Der hohe Stellenwert von RM für Hersteller konkreter Anwendungen wird durch die direkte Adressierung der Entwickler und den Bezug auf bestimmte Updates und Softwareversionen deutlich. Hier werden fundamentale funktionale Probleme wie „cannot log in after update“ oder „new update makes this game unplayable“ zurückgemeldet. Dies bestätigt die besondere Rolle von Reviews, in denen Benutzer den Austausch mit den Verantwortlichen suchen, um Probleme anzumerken. Daher bietet RM vor allem für Entwicklungsteams einzelner Anwendungen die Möglichkeit, schnell auf Nutzerbeschwerden aufmerksam zu werden [22].

Zudem werden gehäuft Softwarefehler und Bugs, Server Probleme und App Crashes sowie ein hoher Batterieverbrauch bemängelt. AR-Apps analysieren die Umwelt durch ein Tracking System und benötigen daher Zugriff auf verschiedene Sensordaten. Die Erhebung und Verarbeitung dieser Daten in Echtzeit stellen hohe Anforderungen sowohl an die Hardware der mobilen Geräte, als auch an die Serverleistung [14]. Daher sollten diese Problemfelder als AR-bezogene Faktoren interpretiert werden.

Eine unmittelbare Relevanz für die Umsetzung von AR zeigt sich in der Beanstandung von fehlendem oder fehlerhaftem Tracking; beispielsweise wird „i think there is something wrong with the compass“ zurückgemeldet. Wie oben erläutert ist das Tracking der Umwelt zentrale Voraussetzung für die erfolgreiche Umsetzung eines AR Erlebnisses. Als fundamentales Problem wird zudem ein fehlender oder funktionsloser AR Modus bemängelt, häufig in Zusammenhang mit Kompatibilitätsproblemen bestimmter Geräte oder Betriebssystemversionen: „AR doesn't work for the 7“.

Nach den Erkenntnissen dieses Topic Models liegen die Quellen von Unzufriedenheit bei Nutzern von AR-Apps vor allem in technischen Grundlagen.

4.3 Topic Modeling in positiver Untersuchungsgruppe

Nach der Filterung der Stichprobe anhand des Sentiment Scores und Ratings stehen noch 9.149 Reviews mit positiven Nutzereindrücken für die Untersuchung zur Verfügung. Diese Auswahl wird von Apps dominiert, die die Erkundung des Nacht- und Sternenhimmels thematisieren: Reviews zu Apps in diesem Themenbereich machen insgesamt 3.942 der Einträge in der Untersuchungsgruppe aus. Auch hier wird die Interpretation des Topic Models soweit wie möglich unabhängig von konkreten App-Inhalten durchgeführt. Die Darstellung der Ergebnisse basiert auf der Interpretation eines 20 Topics umfassenden Topic Models, das in Tabelle 4 im Anhang dargestellt ist. Darin werden erneut sinnvolle Top Words und assoziierte Reviewinhalte zur inhaltlichen Einordnung berücksichtigt.

Analog zur negativen Untersuchungsgruppe werden hier als AR-unabhängige Faktoren ein ausgewogenes Preis-Leistungs-Verhältnis gelobt und eine allgemeine Zufriedenheit und Empfehlung ausgesprochen: „Totally worth it!“, „I would advise this to all my friends“. Ebenso beziehen sich auch die positiven Rückmeldungen gehäuft auf bestimmte Updates und richten sich direkt an die Anwendungsentwickler, was sich in Reviewinhalten wie „The addition of the X-Ray vision improves this already brilliant app“ zeigt. Dabei werden sogar konkrete Verbesserungs- und Erweiterungsvorschläge gemacht, die in der Regel durch „Please add...“ oder „Can you add...“ eingeleitet

werden. Dies macht erneut den hohen möglichen Nutzen von RM als Evaluationstool von spezifischen Apps deutlich.

Die Originalität und der Unterhaltungswert der Apps werden wiederholt positiv zurückgemeldet, z.B. mit „fun and addicting“ und „This is probably the most unique app in the App Store“. Während der reine Spaßfaktor keinen deutlichen AR-Bezug aufweist, kann die Neuartigkeit der Umsetzung von App-Inhalten mithilfe der AR-Technologie einen positiven Einfluss auf die wahrgenommene Originalität bedeuten. Besonders häufig wird zudem die erfolgreiche Umsetzung von Bildungsinhalten in positiven Nutzereindrücken festgehalten, z.B. „This app is making kids love astronomy“ und „all the information is very clear and to the point“. Dies könnte auf den hohen Anteil an Apps mit Bildungsinhalten zum Thema Himmel in der untersuchten Reviewgruppe zurückzuführen sein. Allerdings gehören edukative Anwendungen zu den besonders erfolgversprechenden Kategorien von AR-Apps, zu denen bereits weiterführende Studien zum Nutzerverhalten durchgeführt wurden [4]. Die besondere Eignung und fortgeschrittene Gestaltungsrichtlinien von AR für die Umsetzung von Lernumgebungen kann somit auch zur Erklärung der gehäuften positiven Reviews in diesem Bereich beitragen. Dies gibt Anlass dazu, experimentelle Nutzerdaten auch für andere Anwendungsbereiche zu erheben, um die Nutzererfahrungen auch hier anhand von evidenzbasierten, auf AR-Umgebungen zugeschnittenen Gestaltungsrichtlinien verbessern zu können.

Zentraler Inhalt positiver Rückmeldungen mit eindeutigem Bezug zu AR sind das Interface und die Benutzerfreundlichkeit, was die besonderen Möglichkeiten der AR-Technologie für die Implementierung intuitiver Schnittstellen unterstreicht. Reviews beziehen sich mit Aussagen wie „The way this all was designed and implemented seems incredible“ und „amazingly accurate and intuitive“ auf die grundsätzlichen Stärken der Interfaces. Darüberhinaus werden aber auch multimodale Ansätze gelobt („love this app and also the music“) und trotz spezifischem Bezug zur App „Pokémon Go“ wird der Wunsch nach lokaler Interaktion mit anderen Nutzern deutlich, was die Implementierung kollaborativer AR-Interfaces bekräftigt. Auch der AR-Effekt an sich wird von Nutzern positiv zurückgemeldet. Assoziierte Reviewinhalte wie „This app is my (...) window on the real sky“ und „makes pics come alive!“ zeugen von einer erfolgreichen Umsetzung der „magic lens“ nach Gervautz und Schmalstieg [36]. Zusammenfassend legt dieses Topic Model die Ausnutzung der vielfältigen UI-Gestaltungsmöglichkeiten als zentralen Adoptionstreiber dieser Technologie nahe.

5 Diskussion

Die Topic Models decken Quellen von Lob und Kritik der Nutzer von AR-Apps auf und lassen damit Rückschlüsse zu den Erfolgsfaktoren von AR-Apps zu. Das negative Topic Model verdeutlicht die wiederholte Beanstandung von technischen Problemen: die Umsetzung des Trackings, sowie Verfügbarkeits- und Kompatibilitätsprobleme. Hingegen werden Aspekte der Benutzerschnittstelle nicht kritisiert. Im Einklang damit heben die Ergebnisse des positiven Topic Models die Implementierung eines benutzerfreundlichen Interface als wesentlichen Adoptionstreiber hervor. Die

Ausschöpfung der Interaktions- und Darstellungsmöglichkeiten der AR-Technologie kann somit eine vorteilhafte Mensch-Computer-Schnittstelle erreichen.

Die Durchführung der vorgestellten Untersuchung ist mit Limitationen verbunden. Das Format von App Store Reviews geht mit einer geringen inhaltlichen Qualität der Datengrundlage einher, die in der Verarbeitung und Analyse kompensiert werden muss, sodass trotzdem aufschlussreiche Erkenntnisse gewonnen werden können [22], [37]. Im Hinblick auf die inhaltliche Datenqualität könnten die Untersuchung ergänzt werden, indem die bewertete Nützlichkeit der Reviews berücksichtigt wird. So könnten Reviews, die von mindestens fünf anderen Nutzern als „hilfreich“ bewertet wurden, gesondert ausgewertet werden. Darüber hinaus könnten abwägende Reviews mit mittleren Bewertungen und ohne Zuordnung zu den polaren Untersuchungsgruppen konstruktive Kritik enthalten und sollten daher ebenfalls betrachtet werden.

Wie oben bereits erwähnt, bestätigt die Untersuchung einen hohen Nutzen von RM einzelner Apps für Anwendungsentwickler. Damit können Trends in den Nutzerrückmeldungen schnell erkannt und entsprechende Anpassungen umgesetzt werden. Im Rahmen dieser Untersuchung wurde jedoch angestrebt, allgemeine Schlussfolgerungen in Bezug auf bestimmte App-Eigenschaften zu ziehen. Um dies erfolgreich umzusetzen, hat ein ausgeglichener Datensatz einen hohen Stellenwert. Im hier untersuchten Fall der AR-Apps stehen zum Zeitpunkt der Untersuchung jedoch noch nicht genug Reviews in vielfältigen AR-App-Kategorien zur Verfügung, sodass einzelne Anwendungen wie „Pokémon Go“ und „Sky Guide“ die Analyseergebnisse dominieren. Sobald eine ausreichende Datengrundlage besteht, könnten weiterführende Untersuchungen eine Aufgliederung der Analyse nach den oben angegebenen Kategorien von AR-Apps vornehmen. Darüber hinaus adressieren die identifizierten Faktoren nicht nur spezifische Eigenschaften von AR, sondern auch allgemeine Aspekte, die es in der App-Entwicklung zu berücksichtigen gilt.

Auf methodischer Ebene würde die Interpretierbarkeit der Ergebnisse von einer SA auf dem Level von Aspekten profitieren, sodass verschiedene inhaltliche Bestandteile eines Reviews einzeln klassifiziert werden. Zudem würde die Anlage und Verwendung eines domänenspezifischen Training Sets die Treffsicherheit der Einteilung erhöhen. Das Topic Modeling wurde in dieser Untersuchung mittels eines klassischen LDA Verfahrens durchgeführt, da in der gewählten Data Analyse-Umgebung kein kostenfreies hierarchisches LDA Verfahren verfügbar war. Damit könnten allerdings die Trennschärfe zwischen den Topics und die allgemeine Interpretierbarkeit des Topic Models erhöht werden. Aufgrund der grundsätzlich geltenden eingeschränkten inhaltlichen Qualität von App Reviews und methodischer Einschränkungen kann ein RM in keinem Fall die Ergebnisse einer wissenschaftlichen Nutzerstudie ersetzen.

6 Fazit und Ausblick

Durch die Durchführung eines RM wurden Rezensionen über AR-Apps analysiert. Die Untersuchung trägt mit Erkenntnissen zu den Erfolgsfaktoren von AR-Apps zur Erweiterung der Wissensbasis im Bereich der Applikationsentwicklung bei. Durch die Auswertung von Nutzerbewertungen begegnet diese Untersuchung einer aktuellen

Forschungslücke in der Betrachtung der User Experience von AR-Technologien [38]. Mithilfe von SA wurden Topic Models zu positiven und negativen Benutzerrückmeldungen generiert. Dabei werden vor allem Defizite in den technischen Grundlagen wie dem Tracking System und der Gerätekompatibilität als Entwicklungsfeld identifiziert. Eindeutiger Adoptionstreiber von AR-Apps ist die Umsetzung eines intuitiven Interface mithilfe der AR-Technologie. Die erfolgreiche Anwendung dieses Vorgehens schließt an den von Panichella et al [39] skizzierten Forschungsbedarf zum Einsatz von Topic Modeling in der Reviewauswertung an.

Die Analyse und Auswertung wurde von Wissenschaftlern aus Sicht der App-Entwicklung vorgenommen. Die vorausgehende Erfahrung in der Entwicklung eigener Applikationen stärkt die Validität der Ergebnisse (vgl. [37]). Die identifizierten Erfolgsfaktoren tragen zur Wissensbasis der Entwicklung von AR-Apps bei und sind damit von wissenschaftlicher und praktischer Relevanz. Die gewonnenen Erkenntnisse tragen Implikationen zur Erforschung des Nutzererlebnisses von AR-Applikationen bei (vgl. [40]). Weiterhin unterstützen die vorgestellten Erkenntnisse die erfolgreiche Optimierung und Neuentwicklung von AR-Applikationen.

Die identifizierten Faktoren adressieren auch Gestaltungsempfehlungen, die nicht AR-spezifisch sind. Dies impliziert die weiterhin gegebene Relevanz grundsätzlicher Charakteristika der App-Entwicklung für AR-Anwendungen. Im Einklang mit Forderungen der Wissenschaft sollte darauf aufbauend eine Erweiterung und Adaption vorhandener Gestaltungsrichtlinien für spezifische Technologien vorgenommen werden (vgl. [41]). Die spezifischen AR-Faktoren liefern dazu wertvolle Implikationen für die Entwicklung einer Gestaltungstheorie für AR-Applikationen. Die Ergebnisse können unter anderem im Rahmen von Design Science Research im Bereich der Entwicklung von AR-Anwendungen genutzt und ausgebaut werden [42], [43]. Die Erkenntnisse können durch eine Erweiterung der Analyse auf Nutzerreviews von weiteren Plattformen und neuerscheinenden Applikationen ausgebaut werden. Weiterer Forschungsbedarf besteht hinsichtlich der Übertragbarkeit der Erkenntnisse auf AR-Apps für andere Endgeräte, beispielsweise AR-Glasses.

Literatur

1. Apple Inc.: Introducing ARKit: Augmented Reality for iOS.
2. Google LLC: Introducing ARCore: Augmented Reality at Android Scale (GDD Europe '17).
3. Sutherland, I.E.: The ultimate display. In: Proceedings of the Congress of the International Federation of Information Processing (IFIP). pp. 506–508 (1965).
4. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review. Twenty-third Am. Conf. Inf. Syst. 1–10 (2017).
5. Irshad, S., Rambli, D.R.A.: Advances in Mobile Augmented Reality from User Experience Perspective: A Review of Studies. In: Advances in Visual Informatics: 5th International Visual Informatics Conference, IVIC 2017, Bangi, Malaysia, November 28--30, 2017, Proceedings. pp. 466–477 (2017).
6. Kourouthanassis, P.E., Boletis, C., Lekakos, G.: Demystifying the design of mobile augmented reality applications. *Multimed. Tools Appl.* 74, 1045–1066 (2015).
7. Dey, A., Billingham, M., Lindeman, R.W., Swan, J.E.: A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Front. Robot. AI.* 5, 1–28 (2018).

8. Shitkova, M., Holler, J., Heide, T., Clever, N., Becker, J.: Towards Usability Guidelines for Mobile Websites and Applications. *Wi.* 1603–1617 (2015).
9. Azuma, R.T.A.: *Survey of Augmented Reality*, (1997).
10. Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. *IEEE Comput. Graph. Appl.* 21, 34–47 (2001).
11. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator Telepresence Technol.* 2351, 282–292 (1994).
12. Olsson, T., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: A user study in the context of shopping centres. *Pers. Ubiquitous Comput.* 17, 287–304 (2013).
13. Kim, K., Hwang, J., Zo, H., Lee, H.: Understanding users' continuance intention toward smartphone augmented reality applications. *Inf. Dev.* 1–14 (2014).
14. Mekni, M., Lemieux, A.: *Augmented Reality : Applications , Challenges and Future Trends.* *Appl. Comput. Sci. anywhere.*, 205–214 (2014).
15. Manuri, F., Sanna, A.: A Survey on Applications of Augmented Reality. *Adv. Comput. Sci.* 5, 18–27 (2016).
16. Billinghurst, M., Clark, A., Lee, G.: A Survey of Augmented Reality. *Found. Trends® Human-Computer Interact.* 8, 73–272 (2015).
17. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. *Multimed. Tools Appl.* 51, 341–377 (2011).
18. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits, and atoms. *Proc. ACM SIGCHI Conf. Hum. factors Comput. Syst.* 234–241 (1997).
19. Zhou, F., Dun, H.B.L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In: *Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008, ISMAR 2008.* pp. 193–202 (2008).
20. Reitmayr, G., Schmalstieg, D.: Mobile collaborative augmented reality. In: *Proceedings - IEEE and ACM International Symposium on Augmented Reality, ISAR 2001.* pp. 114–123 (2001).
21. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. *Found. Trends® Inf. Retr.* Vol. 2, No 1-2 1-135. 2, 1–135 (2008).
22. Genc-Nayebi, N., Abran, A.: A systematic literature review: Opinion mining studies from mobile app store user reviews. *J. Syst. Softw.* 125, 207–219 (2017).
23. Martin, W., Sarro, F., Jia, Y., Zhang, Y., Harman, M.: A Survey of App Store Analysis for Software Engineering. 43, 817–847 (2017).
24. Khan, K., Baharudin, B., Khan, A., Ullah, A.: Mining opinion components from unstructured reviews: A review. *J. King Saud Univ. - Comput. Inf. Sci.* 26, 258–275 (2014).
25. Tsytsarau, M., Palpanas, T.: Survey on mining subjective data on the web. *Data Min. Knowl. Discov.* 24, 478–514 (2012).
26. Blitzer, J., Dredze, M., Pereira, F.: Domain adaptation for sentiment classification. *Annu. Meet. Comput. Linguist.* 45, 440–447 (2007).
27. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Comput. Surv.* 50, 1–33 (2017).
28. Debortoli, S., Müller, O., Junglas, I., Vom Brocke, J.: Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Commun. Assoc. Inf. Syst.* 39, (2016).
29. Rossetti, M., Stella, F., Zanker, M.: Analyzing user reviews in tourism with topic models. *Inf. Technol. Tour.* (2016).
30. Blei, D., Ng, A.: Latent dirichlet allocation. *J. Mach. Learn.* 3, 993–1022 (2003).
31. Blei, D.M.: Probabilistic topic models. *Commun. ACM.* 55, 77–84 (2012).

32. Blei, D., Griffiths, T.L., Jordan, M.I., Tenenbaum, J.B.: Hierarchical topic models and the nested Chinese restaurant process. *Adv. Neural Inf. Process. Syst.* 16, 106 (2004).
33. Crystal, D.: A global language. In: *English in the World*. pp. 151–196 (2013).
34. Bye, K.: Scrape iOS app review data from iTunes.
35. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*. 10, 2200–2204 (2010).
36. Gervautz, M., Schmalstieg, D.: Anywhere interfaces using handheld augmented reality. *Computer (Long Beach, Calif.)*. 45, 26–31 (2012).
37. Chen, N., Lin, J., Hoi, S.C.H., Chen, N., Lin, J., Hoi, S.C.H., Xiao, X., Zhang, B.: AR-Miner : Mining Informative Reviews for Developers from Mobile App Marketplace AR-Miner : Mining Informative Reviews for Developers from Mobile App Marketplace. (2014).
38. Parvinen, P., Hart, A., Griswold, M., Lai, J., Boyer, E.W., Skolnik, A.B., Chai, P.R.: Introduction to Minitrack : Mixed , Augmented and Virtual Reality. 1395–1396 (2018).
39. Panichella, S., Sorbo, A. Di, Guzman, E., Visaggio, C.A., Canfora, G., Gall, H.C.: How Can I Improve My App ? Classifying User Reviews for Software Maintenance and Evolution. *2015 IEEE Int. Conf. Softw. Maint. Evol.* 281–290 (2015).
40. Niemöller, C., Metzger, D., Thomas, O.: Design and Evaluation of a Smart-Glasses-based Service Support System. In: *13. International Conference on Wirtschaftsinformatik (WI)*. pp. 106–120 (2017).
41. Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson (2016).
42. Metzger, D., Niemöller, C., Jannaber, S., Berkemeier, L., Brenning, L., Thomas, O.: The Next Generation: Design and Implementation of a Smart Glasses-based Modelling System. *Enterp. Model. Inf. Syst. Archit. - Int. J. Concept. Model.* [accepted for publication] (2018).
43. Zobel, B., Berkemeier, L., Werning, S., Vogel, J., Ickerott, I., Thomas, O.: Towards a modular reference architecture for smart glasses-based systems in the logistics domain. In: *CEUR Workshop Proceedings*. p. Paper 17 (2018).

Anhang

Tabelle 2. Deskriptive Daten zu gescraptten Reviews und Untersuchungsgruppen

<i>Untersuchte App</i>	<i>Bewertung</i>	<i>Reviews</i>	<i>Reviews in negativer Gruppe</i>	<i>Reviews in positiver Gruppe</i>
Sky Guide	4,9	10.884	23	2.865
Pokémon Go	3,7	86.217	397	433
Egg, Inc.	4,8	21.666	14	1.219
Zombie Gunship Revenant AR	4,7	415	6	56
Star Walk 2 – Night Sky Map	4,8	2.559	17	1.200
CamToPlan PRO	4,6	527	8	215
Tape Measure	4,2	153	9	19
Star Walk 2 Ads+: Sky Guide	4,8	408	10	142

<i>Untersuchte App</i>	<i>Bewertung</i>	<i>Reviews</i>	<i>Reviews in negativer Gruppe</i>	<i>Reviews in positiver Gruppe</i>
Lumyer	4,7	2.898	14	1.710
magicplan	4,7	2.045	38	672
Thomas & Friends Minis	4,5	168	3	32
AirMeasure – AR Tape & Ruler	4,6	95	3	20
Euclidean Lands	4,7	155	7	19
Human Anatomy Atlas 2018	4,9	163	0	56
Stack AR	3,8	472	47	36
Monster Park – AR Dino World	4,5	90	6	21
Heads Up Navigator	2,3	37	3	6
Plotaverse Photo Video Editor	4,5	471	6	90
Inkhunter Try Tattoo Designs	4,7	782	12	329
GIPHY World: AR GIF Stickers	4,6	57	3	9

Tabelle 3. Topic Model zu negativer Untersuchungsgruppe mit 10 Topics

<i>Themenbereich</i>	<i>Inhaltliche Interpretation</i>	<i>Top Words</i>	<i>Assoziierte Reviewinhalte</i>
Preis-Leistungs-Verhältnis, Empfehlung	Allgemeine Unzufriedenheit Preis-Leistungs-Verhältnis	download, delete, useless waste_time, money, waste	„huge regret downloading“, „won't download“, „The game is useless...“ „It's a waste of money“, „This is so bad I do not like this I want my money back“
Direkte Kommunikation	Direkte Entwickler-kommunikation Probleme nach Update Login Probleme nach Update	company, customer, user worst_update, ruin, unplayable update, broken, account	„A shameless cash grab from a company with a poor track record from listening to their customers.“ „New updated makes this game unplayable, not playing until it is fixed“ „Cannot log in after update“

<i>Themenbereich</i>	<i>Inhaltliche Interpretation</i>	<i>Top Words</i>	<i>Assoziierte Reviewinhalte</i>
Technische Probleme	Softwarefehler, Bug	play, freeze, problem	„lag and freeze anytime“, „freezes, hard to use“
	Server Probleme, App Crashes	crash, server, keep_crash	„This game is crashing since i got it...“, „Needs More Servers“
	Batterieverbrauch	bad, battery	„has now become a huge drain on my battery life since the battery saver removal“
	Tracking System, Sensoren	compass, wrong	„I think there is something wrong with the compass in all iphone 6s models“,
	AR Modus, Kompatibilität	won't, work, ar_mode, ar_work	„Very stupid how the AR doesn't work for the 7“, „AR doesn't work, no visual feedback in order to track the surface“

Tabelle 4. Topic Model zu positiver Untersuchungsgruppe mit 20 Topics

<i>Themenbereich</i>	<i>Inhaltliche Interpretation</i>	<i>Top Words</i>	<i>Assoziierte Reviewinhalte</i>
Preis-Leistungs-Verhältnis, Empfehlung	Preis-Leistungs-Verhältnis	worth, purchase, price	„Well worth 1.49... probably more like \$5.00“, „Totally worth it!“
	Preis-Leistungs-Verhältnis, Empfehlung	definitely_worth, thumbs	„Great app, definitely worth the 2.99“, „Definitely recommend this app“
	Allgemeine Zufriedenheit, Empfehlung	amazing, favorite, friends, family	„What an amazing app this is... I have shared it with all my grandkids.“, „I would advise this to all my friends“
Direkte Kommunikation	Verbesserung nach Update	version, feature, update, improve	„The addition of the X-Ray vision improves this already brilliant app“
	Direkte Entwicklerkommunikation	work, keep, update	„Love it. Now they need to keep building on it!“, „I can't wait for the next update! When's it coming out?“
	Direkte Entwicklerkommunikation	please, add, option	„Please add...“, „Can you add...“

<i>Themenbereich</i>	<i>Inhaltliche Interpretation</i>	<i>Top Words</i>	<i>Assoziierte Reviewinhalte</i>
Originalität, Unterhaltung	Spaß	addictive, fun_play	„Fun and addicting“, „I haven't stopped playing for days“
	Originalität	nice, clever, unique	„This is probably the coolest and most unique app in the App Store“
	Originalität und Spaß	time, idea	„play it all the time“, „Great idea“, „the app is definitely ONE OF A KIND!“
Bildungsinhalte	Bildungsinhalte	learn, help, know	„An outstanding app! This really helps me find and learn more about the stars...“
	Bildungsinhalte	learn, tool, kids	„I got it for my daughter because she wants to know more about the universe“, „This app is making kids love astronomy“
	Bildungsinhalte	information, detailed, accurate	„All the information is very clear and to the point“, „Chock full of information at one's fingertips“
	Bildungsinhalte, Interface	beautiful, graphics, informative, educational	„A gorgeous, fluid interface“, „This app is so beautiful and educational!“
Benutzerfreundlichkeit	Benutzerfreundlichkeit, Effektivität	user_friendly, easier	„can make your work that much easier“, „awesome, user friendly“
	Interface	impressive, design, quality	„The way this all was designed and implemented seems incredible.“
	Interface	easy, fun, interface	„Great job on the UI, really easy to use“, „amazingly accurate and intuitive“
	Kollaboratives Interface	trade, gym, team	„I would love to see the day when you can chose a battle with another local user and also being able to trade your Pokémon“
	Multimodales Interface	relaxing, music	„love this app and also music – so relaxing“
Immersion	Immersion	look, see, show	„This app is my only window on the real sky“, „whats better than to look up and map what you see with what's on your phone. Magical“
	Immersion	effects, life, creative	„fun photoapp that makes a picture more alive“, „makes pics come alive! So creative!“