KOSMO

KI-gestützte kollektiv-soziale Moderation von Online-Diskursen (KOSMO)

Verbundprojekt zwischen Liquid Democracy e.V. (Lead), dem Institut für Partizipatives Gestalten und der Heinrich-Heine-Universität Düsseldorf

Förderung durch das Bundesministerium für Bildung und Forschung (BMBF), Progammlinie "KMU innovativ" (FKZ: 01IS19040B)

Gemeinsamer Abschlussbericht zum Verwendungsnachweis

Projektverantwortliche:

Liquid Democracy e.V. (LIQD)	Marie-Kathrin Siemer
Heinrich-Heine-Universität Düsseldorf (HHU)	Dr. Katharina Gerl und Prof. Dr. Marc Ziegele
Institut für Partizipatives Gestalten/ Hörster und Rohr GbR (IPG)	Roland Ronja Wehking

Kontakt: Marie-Kathrin Siemer / Liquid Democracy e.V., Am Sudhaus 2, 12053 Berlin / Tel.: 030-62 984840 / m.siemer@liqd.net

Berlin, den 13.10.2023

Liquid Democracy

INSTITUT FÜR PARTIZIPATIVES GESTALTEN IPG

hhu Heinrich Heine Universität Düsseldorf

KOSMO

Inhaltsverzeichnis

Sachbericht zum Verwendungsnachweis Teil 1

Das trans- und interdisziplinäre Vorhaben „KI-gestützte kollektiv-soziale Moderation von Online-Diskursen“ (KOSMO) ist ein Verbundprojekt zwischen Liquid Democracy e.V. (LIQD, Lead), dem Institut für Partizipatives Gestalten (IPG) und der Heinrich-Heine-Universität Düsseldorf (HHU). Das Projekt wurde für die Dauer von drei Jahre (2020-2023) vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Kennzeichen FKZ: 01IS19040B in der Programmlinie “KMU innovativ” gefördert.

Digitale Plattformen sind wichtige Instrumente für die gemeinsame Entwicklung von Ideen. Sie sind auch Orte des politischen Meinungsaustauschs und der Meinungsbildung. Eingeschränkt werden diese Möglichkeiten zum gesellschaftlichen Diskurs durch Phänomene wie Online-Inzivilität (z.B. Hassrede). Ein Mittel, diesen Phänomenen zu begegnen, ist die Moderation von Online-Diskussionen, die zunehmend auch automatisiert eingesetzt wird, um problematische Inhalte zu identifizieren. Trotz deren Wirksamkeit kommt Moderation jedoch noch nicht flächendeckend zur Anwendung. Motivation und Ausgangspunkt für KOSMO war die Beobachtung, dass Initiator:innen von Online-Diskussionen (Politik und Verwaltung, zivilgesellschaftliche Organisationen, Medienunternehmen) vor drei zentralen Herausforderungen stehen, die der weiteren Verbreitung und Implementation von Diskussionsformaten für die demokratische Beteiligung im Wege stehen:

1. Die Moderation von Online-Diskussionen ist sehr aufwändig.
2. Potentielle Teilnehmende werden durch mangelnde Diskussionsqualität abgeschreckt.
3. Eine Verwertung der Ergebnisse wird durch die aufwändige Auswertung von Online-Diskussionen erschwert.

Ziel des Projekts KOSMO war deshalb die prototypische Entwicklung eines KI-gestützten Assistenzsystems, das die Moderation bei der Qualitätssicherung und Synthese von Online-Diskussionen und Online-Partizipationsverfahren proaktiv unterstützt und sie teilweise automatisiert. Folgende Schwerpunkte sollten bei der Entwicklung und Erprobung gesetzt werden:

1. Die Reduzierung des Moderationsaufwands, indem eine KI-gestützte automatisierte umfassende Analyse der Qualität von Beiträgen stattfindet.
2. Die proaktive Unterstützung von Moderator:innen bei der Beantwortung von problematischen und wenig argumentativen Beiträgen.
3. Eine Erhöhung der Beteiligungsmotivation von bislang passiven Diskussionsbeobachter:innen.
4. Eine bessere Verwertbarkeit der Ergebnisse, indem durch Mechanismen der Gamifizierung die Zusammenfassung von Beiträgen durch Teilnehmende unterstützt wird.
5. Eine modulare Umsetzung eines Prototyps unter Open-Source-Lizenz, der von verschiedenen Zielgruppen praktisch erprobt wird.

Die Durchführung des Projekts erfolgte in iterativen Prozessschritten. Hierdurch wurde berücksichtigt, dass es sich bei der Software- und KI-Entwicklung um anspruchsvolle und hochdynamische Vorgänge handelt, welche ständiger Anpassung an technische Innovationen und Rejustierung auf der Basis von Nutzer:innenfeedback benötigen. Dies wurde in der Projektkonzeption berücksichtigt, indem zwischen den einzelnen Entwicklungsphasen sogenannte **Praxistests** zur praktischen Erprobung und Evaluation des Prototyps mit assoziierten Partnern eingeplant wurden (s. Abb. 1).

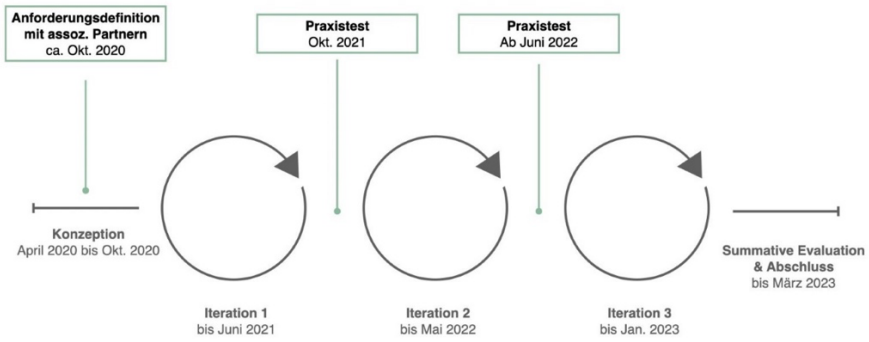

Abbildung 1: Schematischer Projektverlauf

KOSMO schloss an drei wissenschaftliche Forschungsschwerpunkte an: 1) Die Erforschung der Qualität von Online-Diskussionen, 2) die Wirkungen von Moderation und Zusammenfassungen dieser Diskussionen und 3) die automatisierte Detektion von qualitativ hoch- und minderwertigen Beiträgen mittels Künstlicher Intelligenz (v.a. Maschinelles Lernen).

Das Ergebnis von KOSMO ist die Entwicklung und Veröffentlichung eines KI-gestützten Assistenzsystems als Prototypen, der unter https://kosmo.liqd.net frei zugänglich verfügbar gemacht wurde. Die wesentlichen sieben Erfolgsmerkmale dieses Prototyps sind:

1. Entwicklung und Erprobung von Algorithmen für die automatisierte Detektion von qualitativ minder- und hochwertigen Diskussionsbeiträgen in deutscher Sprache.
2. Reduzierung des Moderationsaufwandes durch die automatisierte Analyse der Beiträge durch eine KI, die inzivile (qualitativ minderwertige) Beiträge sowie deliberativ-bereichernde (qualitativ-hochwertige) Beiträge sowie Kommentare mit Tatsachenbehauptungen erkennt.
3. Proaktive Unterstützung von Moderator:innen bei der Beantwortung von problematischen Beiträgen durch die Einbindung eines nutzer:innenzentrierten Designs, der Umsetzung von Elementen der Meaningful Gamification und der Entwicklung von Workshops zur Schulung und zum Empowerment von Moderator:innen.
4. Erhöhung der Beteiligungsmotivation durch Einbezug und Bindung der Teilnehmenden in die “Storyline” der Plattform mit Elementen der Meaningful Gamification, indem insbesondere die Beziehung zwischen Moderation und Teilnehmenden an der Diskussion gestärkt wurde.
5. Anschließende Forschungsfragen zur Weiterentwicklung der Software und der KI zur Synthese von Beiträgen von Teilnehmenden und damit der besseren Verwertbarkeit der Kommentare sowie eines einfacheren Einstiegs in die Diskussion für neue Diskussionsteilnehmende.
6. Erprobung des Prototypens in zwei Praxistests am Ende der jeweiligen Iterationen.
7. Beitrag zur wissenschaftlichen Grundlagenforschung und zur anwendungs- und gestaltungsorientierten Forschung zur Rolle von KI in demokratischen Prozessen.

Sachbericht zum Verwendungsnachweis Teil 2

1. Einleitung

Ziel des Projekts KOSMO war die Entwicklung des Prototyps eines KI-gestützten Assistenzsystems, das die Moderation bei der Qualitätssicherung und der Synthese von Online-Diskussionen und Online-Partizipationsverfahren proaktiv unterstützt und sie teilweise automatisiert. Im Folgenden stellen die Konsortialpartner ihre Projektarbeit entlang der Arbeitspakete vor. Dabei ist klar markiert, welcher Konsortialpartner – Liquid Democracy e.V. (LIQD), Heinrich-Heine-Universität (HHU), Institut für Partizipatives Gestalten (IPG) - für welches Ergebnis verantwortlich ist, auch wenn mehrere Konsortialpartner innerhalb eines Arbeitspakets tätig waren. Schließlich erfolgt eine Darstellung der Verwertbarkeit der Ergebnisse sowie der erfolgten und geplanten Veröffentlichungen, Publikationen und Veranstaltungen.

2. Aufschlüsselung der Projektarbeit nach Arbeitspaketen

AP 1 Anforderungsanalyse und -definition

Ziel des Arbeitspaketes „Anforderungsanalyse und -definition“ war es, einen Anforderungskatalog für KOSMO basierend auf aktuellen wissenschaftlichen Studien, bereits existierenden Systemen und Anforderungen unserer Praxispartner zu erstellen. Dazu erfolgte eine wissenschaftliche Aufarbeitung der derzeitigen theoretischen und empirischen Forschungsergebnisse im Bereich der Online-Deliberation durch die HHU. Hier wurden Indikatoren herausgearbeitet, die als Gradmesser der Qualität von Online-Deliberation dienen können. Basierend auf diesen Ergebnissen erfolgte eine Potenzialanalyse durch LIQD, um festzustellen, welche (KI-gestützten) Software-Lösungen zur Moderation von Online-Diskussionen bereits zur Verfügung stehen. Im Verhältnis zu den Forschungsergebnissen und den bereits vorhandenen Software-Lösungen erschloss sich, dass es eine Marktlücke für ein Moderationstool gibt, das die deliberative Qualität in partizipativ-politischen Online-Diskussionen über die Vermeidung von Inzivilität hinaus und die Förderung von konstruktiven Beiträgen semi-automatisiert unterstützt. Diese Potenzialanalyse inklusive des Literaturberichts war – wie in der Vorhabenbeschreibung angedacht – die Grundlage für die weitere Planung. So führten wir im Anschluss daran interne Workshops unter den Konsortialpartnern sowie Workshops mit Vertreter:innen unserer Praxispartner durch – v.a. Moderator:innen von Online-Partizipationsverfahren und von Kommentarspalten auf Nachrichtenmedien – um zu überprüfen, ob sich unsere Erkenntnisse mit einem tatsächlichen Bedarf der potenziellen Zielgruppen von KOSMO überschneiden. Die Konzeption, Organisation und Moderation der Workshops wurde durch das IPG übernommen. Die wissenschaftliche Dokumentation, Auswertung und Präsentation der Ergebnisse der Workshops erfolgte durch die HHU. Auf Basis des Zielgruppen-Workshops hat LIQD dann einen priorisierten Anforderungskatalog erstellt, in dem die vorrangigen Bedarfe der Zielgruppen mit konkreten Feature-Ideen von KOSMO gelistet waren. Zusammen im Konsortium haben wir dann entschieden, welche Punkte wir als erstes umsetzen wollen. Die Erkenntnisse haben wir im “Outro AP1” vorgestellt. Sie dienten als Basis für das weitere Vorgehen in den Bereichen Software-Entwicklung und Künstliche Intelligenz (insbesondere AP 2 und AP 3). Die wichtigsten Positionen des zahlenmäßigen Nachweises in AP 1 sind die Beschäftigung des Projektmanagements, die zur Durchführung notwendig waren sowie die Workshopdurchführungen.

Zusammenfassend wurden alle in AP 1 geplanten Ergebnisse erreicht: (1) Literaturbericht bestehend aus Literaturbericht und Potenzialanalyse und (2) Anforderungskatalog für KOSMO.

AP 2 Entwicklung und Training der KI-Algorithmen

Ziel des Arbeitspaketes 2 (AP 2) war die Entwicklung von Algorithmen für das Moderationstool KOSMO, die aus den folgenden Teilschritten bestand: (1) Generierung vielfältiger und qualitativ hochwertiger Trainingsdaten; (2) Anlernen von state-of-the-art Algorithmen zur Erkennung von inzivilen und deliberativen Nutzerbeiträgen; (3) Implementierung der Algorithmen in KOSMO über eine API (in enger Zusammenarbeit mit LIQD); (4) Evaluation der Algorithmen (in Zusammenarbeit mit AP 6 der HHU). Als wichtigste Positionen des zahlenmäßigen Nachweises in AP 2 sind die Beschäftigung der wissenschaftlichen Mitarbeiterin, die Ausgaben für manuelle Codierarbeiten durch wissenschaftliche Hilfskräfte sowie die Ausgaben für eine umfangreiche Annotation von Trainingsdaten durch Clickworker zu nennen. Diese Ausgaben waren notwendig, um die im Arbeitspaket anfallenden Aufgaben zu übernehmen. Diese umfassten die folgenden Arbeitsschritte: A) Die Aufbereitung bestehender Datensätze, die für das Training der Algorithmen relevant waren. B) Die Kommunikation mit Projektpartnern zur Bereitstellung von Trainingsdaten. C) Die Konzeption von Codebüchern zur Codierung/Annotation der Trainingsdaten durch wissenschaftliche Hilfskräfte und durch Clickworker. D) Die Durchführung von Codierschulungen und Betreuung der Hilfskräfte bei der Codierung. E) Die Einarbeitung in Forschungsstand zu KI-basierter Klassifizierung von Nutzerkommentaren. F) Die Organisation und Ausrichtung des Hackathons GermEval 2021 (in Kooperation mit AP 6 der HHU), in dessen Rahmen state-of-the-art Algorithmen zur Klassifikation von Deliberation und Inzivilität in Online-Diskussionen entwickelt, evaluiert und publiziert wurden. G) Die Programmierung sowie das Training und Testing der Algorithmen für verschiedene Kommentarkategorien. H) Die Einarbeitung in Software-Entwicklung zur Erstellung der API (in Kooperation mit LIQD). I) Die Konzeption und Programmierung der API zur Kommunikation mit der Software KOSMO sowie deren Weiterentwicklung im Laufe der drei Iterationen (in enger Zusammenarbeit von HHU und LIQD). J) Die Konzeption, Durchführung und Auswertung der Begleitstudie zur Crowdannotation (in Kooperation mit AP 6 der HHU).

Insgesamt konnten nahezu alle Arbeiten im Arbeitspaket mit den avisierten Ergebnissen abgeschlossen werden: (1) Eine umfangreiche Datenbank mit (2) sowohl von wissenschaftlichen Hilfskräften als auch von Crowdworkern annotierten Trainingsdaten liegt vor und (3) die Algorithmen zur Detektion von deliberativen und inzivilen Kommentaren wurden getestet, implementiert, iterativ überarbeitet und schließlich als maschinenlesbarer Code Open Source verfügbar gemacht. Schließlich wurden die Ergebnisse auf (5) zahlreichen Fachtagungen und in wissenschaftlichen Publikationen festgehalten (siehe AP 7). Lediglich (4) das Whitepaper, das die Funktionsweise und Performance-Werte der Algorithmen im Überblick enthält, wird aktuell noch finalisiert, um die letzte Iteration abzubilden.

AP 3 Software-Konzeption und -Entwicklung

Ziel des AP 3 „Software-Konzeption und -Entwicklung“ war die Konzeption und Entwicklung eines Prototyps des webbasierten Assistenzsystems KOSMO. Basierend auf der Open Source Software adhocracy wurde ein Moderationstool entwickelt, das in Praxistests getestet wurde und in das Erkenntnisse und Ergebnisse aus den AP 2 und 4 integriert wurden. Adhocracy ist eine Beteiligungssoftware, die verschiedene Module zur Verfügung stellt, um einen Beteiligungsprozess transparent darzustellen und Teilnehmende in unterschiedliche Momente der Beteiligung einzubinden. Für KOSMO entwickelten wir adhocracy mit dem Fokus auf Moderation und Interaktion von Teilnehmenden und Moderation weiter. Die Weiterentwicklung erfolgte in drei Iterationen, nach jedem Umsetzungsschritt wurden die

Funktionalität und der Bedarf in einem Praxistest überprüft. Die Umsetzung orientierte sich an dem priorisierten Anforderungskatalog aus AP 1, der nach jeder der drei Iterationen innerhalb des Konsortiums und nach Einbezug der Erkenntnisse aus den Praxistests angepasst und erneut gemeinschaftlich priorisiert wurde. Dazu fanden in den jeweiligen Iterationen wiederholt interne Workshops zur Konzeption und Priorisierung von User Stories statt. Die gesamte Software-Konzeption und Umsetzung der Software erfolgte mit der agilen Projektmanagement- und Software-Entwicklungs-Methode SCRUM.

Auf Basis des technischen Konzeptes erfolgte die Gestaltung und Implementierung des Frontends sowie die technische Implementierung des Backends durch LIQD iterativ in Anlehnung an die Iterationen. Dazu erstellte LIQD Wireframes und Designs, die in internen Meetings innerhalb des SCRUM-Prozesses von den Konsortialpartner:innen, dem Frontend- und dem Backend-Team besprochen und anschließend umgesetzt wurden. Der Fokus der ersten Iteration war das Aufsetzen der Plattform sowie die Entwicklung eines Nutzer:innen-Dashboards, in dem sich Nutzer:innen und die Rolle der Moderation über anstehende Beteiligungsprojekte individuell informieren können. Eine wichtige Umsetzung in der ersten Iteration war außerdem die Implementierung und das Testen der KI in enger Abstimmung zwischen LIQD und der HHU. Eine vorläufige Entscheidung und Umsetzung einer ersten Version zur Darstellung der KI-Kategorien im Moderations-Dashboard wurde innerhalb des Konsortiums besprochen und anschließend umgesetzt und getestet. In der zweiten Iteration lag der Fokus deshalb auf der Verbesserung der Darstellung der von der KI und anderen Teilnehmenden gemeldeten Kommentare. Dazu arbeitete LIQD in der zweiten Iteration eng mit AP 2 (HHU) an einer besseren Usability für die Moderation, um Moderations-Abläufe und die Übersicht der gemeldeten Kommentare zu verbessern. Hier war die Herausforderung, dass die gemeldeten Kommentare unterschiedliche Aktionen verlangen, je nachdem, ob sie von der Moderation als deliberativer Kommentar (hohe Qualität) oder als inziviler Kommentar (niedrige Qualität) identifiziert worden sind. So würde ein Kommentar mit hoher Qualität eher hervorgehoben werden, wohingegen ein inziviler Kommentar verborgen oder kritisch beantwortet werden würde. Diese Herausforderung spiegelte sich auch in der dritten Iteration wider, die sich schließlich auf die Optimierung der Kommunikation zwischen KI, Moderation und Nutzer:innen in enger Zusammenarbeit mit AP 4 (IPG) fokussierte. Es wurden das Styling verbessert, die Meldungen von KI, Nutzer:innen und die möglichen Aktionen der Moderation in Bezug auf die gemeldeten Kommentare optimiert und die Darstellung der KI sowie das Notification-System überarbeitet.

Dabei veröffentlichte LIQD während aller Iterationen den Quellcode mit einer Open-Source-Lizenz (GNU Affero General Public License v3.0) sowie dessen Dokumentation im Code Repository (siehe: https://github.com/liqd/a4-kosmo). Für das stetige manuelle und automatische Testing setzte LIQD ein Entwicklungs-, ein Test- und ein Produktivsystem auf. Auf alle Systeme wurde regelmäßig die aktuell entwickelte Version von KOSMO deployed. Das Produktivsystem mit dem fertigen und einsetzbaren Prototyp ist zu finden unter: kosmo.liqd.net

Der finale Prototyp umfasst eine multi-modulare Beteiligungsplattform mit einem Dashboard für Nutzer:innen und einem weiteren Dashboard zur Übersicht aller gemeldeten Kommentare für die Rolle der Moderation. In dem Dashboard können individuelle Aktivitäten und gefolgte Projekte und Organisationen angezeigt werden. Außerdem gibt es auch den Hinweis auf Hintergrundinformationen zur Plattform und zur Nutzung von KI auf der Seite. Die Rolle der Moderation hat eine Übersicht über alle zu moderierenden Projekte mit einer Anzeige, wie viele Kommentare es bereits in dem Projekt gab und wie viele gemeldete Kommentare unbearbeitet sind. So kann die Moderation bereits bei den Projekten priorisieren, in welchem Beteiligungsprojekt schnelle Aktivität der Moderation nötig ist. Die Auflistung der gemeldeten Kommentare im Dashboard des jeweiligen Beteiligungsprojektes kann nach

Bedarf sortiert und gefiltert werden. Die KI übernimmt dabei eine essentielle Aufgabe in der Vorsortierung: Sie markiert die Kommentare automatisiert entlang der in AP 2 definierten Kategorien, sodass die Moderation hier schnell schädliche oder konstruktive Kommentare sowie Tatsachenbehauptungen identifizieren kann. Dabei ist eindeutig markiert, welche Kommentare von der KI oder anderen Teilnehmenden gemeldet worden sind. Außerdem ist die Liste der gemeldeten und bearbeiteten Kommentare stets aktuell, sodass auch mehrere Moderator:innen gleichzeitig Kommentare bearbeiten können. Die gemeldeten Kommentare können direkt im Dashboard mit Effekt auf die Diskussion im Nutzer:innen-Interface verborgen oder hervorgehoben werden. Die Moderation kann außerdem direkt auf den gemeldeten Kommentar mit einem weiteren Kommentar reagieren. Eine Verlinkung direkt in die Diskussion ist gegeben, sodass die Moderation bei Bedarf den Kommentar im Kontext der Diskussion lesen kann. Im Nutzer:innen-Interface werden die hervorgehobenen Kommentare farblich und symbolisch von den anderen Kommentaren abgehoben, die Kommentare der Moderation haben eine prominente Position direkt unter dem jeweiligen Kommentar, um eine direkte Reaktion der Moderation im Falle von Inzivilität aber auch im Falle besonders guter Kommentare oder noch zu überprüfenden Fakten darstellen zu können.

Die wichtigsten Positionen im zahlenmäßigen Nachweis von Liquid Democracy sind die Personalkosten der Entwickler:innen, die die Umsetzung des Prototypens sicherstellten, sowie des Designers und der Projektmanager:innen, die signifikanten Anteil an der co-kreativen Konzeption und an den Absprachen zwischen internem Team von LIQD und den Konsortialpartner:innen hatten. Insgesamt wurden auch hier alle Arbeitsschritte und Meilensteine erreicht: (1) Katalog von User Stories und Dokumentation nicht-funktionaler Anforderungen, (2) Wireframes, UI-Designs für Iteration 1, Anpassungen des UI-Designs für Iteration 2 und 3 sowie die Frontend-Programmierung (CSS, HTML, JavaScript) Iteration 1–3, (3) Technisches Umsetzungskonzept zur Einbindung der KI-Komponenten aus AP2; Backend- Programmierung für Iteration 1–3 des KOSMO-Prototyps, (4) Test/-Entwicklungssystem, Produktivsystem, Deployment der Iterationen 1–3 auf dem Produktivsystem und (5) Code Repository, Dokumentation der Installation in Form einer Readme-Datei, Veröffentlichung des Quellcodes von Iteration 2 und 3 auf dem Repository.

AP 4 Gamifizierung und User-Centred Design

Ziel des AP4 Gamifizierung und User-Centred Design war es, durch die Implementierung von Elementen der „Meaningful Gamification“ die Grundlage für die spätere Nutzerakzeptanz der Software zu legen. So sollte die Nutzbarkeit der Plattform verbessert und Anforderungen sowie Empfehlungen von Gamifizierungs-Mechanismen zur Synthese von Beiträgen erstellt werden.

Wie im Antrag beschrieben, wurden bekannte Mechanismen wie Ranglisten, Abzeichen und Diagramme evaluiert und mit den Zielen der Plattform in Beziehung gesetzt. Dies geschah unter anderem im Rahmen zweier vom IPG organisierter, virtuell durchgeführter Workshops mit Projekt- und Konsortialpartner:innen im Jahr 2021. Am Workshop nahmen auch Mitglieder der HHU und LIQD teil. Dabei wurde unter anderem durch die Zuhilfenahme von Userstories bestätigt, dass eine Implementierung von Konzepten der „Meaningful Gamification“ strukturell deutlich tiefer als die angesprochenen Mechanismen gehen müssen. Darauf aufbauend wurde der Fokus der Gamifizierung im Laufe des Projektes auf die generelle Nutzbarkeit der Plattform erweitert. AP 4 war in der ursprünglichen Vorhabenbeschreibung noch insbesondere auf die Synthese von Nutzer:innenbeiträgen fokussiert gewesen. Als zusätzliches Framework wurde neben Nicholsons „Meaningful Gamification“ noch Yu-Kai Chows „Octalysis“ Framework und die dort beschriebene „Spielerreise“ eingebunden. Diese erstreckt sich

über die Bereiche Discovery, Onboarding, Scaffolding und Endgame und wurde für die Nutzer:innengruppen: „Moderator:innen“ und „Diskussionsteilnehmende“, sowie „Übergeordnet“ nachgezeichnet. Es formte sich das Verständnis, dass die Plattform mit einem besonderen Fokus auf die Beziehung zwischen Moderator:innen, KI und Nutzer:innen zu gestalten sei. Schnelle, direkte Kommunikationswege, die es erlauben, nuanciert miteinander zu kommunizieren, Feedback zu geben und auf Beiträge zu reagieren, erscheinen für KOSMO von deutlich höherer Relevanz als die üblichen „Points, Badges & Leaderboards“ aus dem Gamifizierungsbereich.

Zusätzlich war es wichtig, ein klares, gut zugängliches Onboarding zu gestalten, sodass alle Funktionen der Plattform sowohl von Moderation als auch Nutzer:innen schnell wahrgenommen werden konnten. Diese Erkenntnisse flossen direkt in die Software-Entwicklung und Software-Konzeption ein, und LIQD setzte in der dritten Iteration Features um, die direkt auf die Verbesserung der Kommunikation zwischen Moderation und Nutzer:innen einzahlte, unter anderem die Darstellung eines Dummy-Projektes im Teilnehmenden-Dashboard, die Verbesserung des Registrierungs- und Login-Prozesses inklusive neuer Willkommensmail, die technische Umsetzung und inhaltliche Befüllung einer statischen Informationsseite zu KOSMO und die Verbesserung und erweiterte Darstellung des Activity-Feeds im Teilnehmenden Dashboard.

Für die nutzer:innengenerierte Synthese wurden in einem Workshop im Mai 2022 zunächst aus Nutzer:innenperspektive mögliche Arten der Synthese gestaltet. An diesem internen Workshop unter der Moderation des IPG nahmen Mitglieder von HHU und LIQD teil. Ergebnis waren einerseits die Nutzung von Hashtags, um schnell einen Überblick über die relevanten Themen einer Diskussion zu gewinnen. Ebenso könnten Synthesen einzelner Beiträge genutzt werden, um Nutzer:innen den Einstieg in die Diskussion zu erleichtern. Diese beiden Varianten wurden in einem AP-4-internen Usertesting im März 2023 mit Hilfe eines von LIQD entwickelten und mehrfach iterierten Clickdummies getestet. Die Teilnehmenden lasen dafür zunächst einen Text und Kommentar und konnten anschließend Synthesen und Hashtags nutzen, um zu bewerten, inwiefern diese dabei helfen einen schnellen Einstieg in die Diskussion zu gewinnen. Die semi-automatisierte Synthese von Kommentaren konnte im derzeitigen Moderationstool nicht umgesetzt werden, da sich hier größere konzeptuelle Fragen sowie Fragen der technischen Möglichkeit der KI gestellt haben. Es ist angestrebt, die Ergebnisse des Praxistests in einem Folgeprojekt umzusetzen. Ebenso wie im Gesamtprojekt ist im AP Gamifizierung die iterative, kokreative Zusammenarbeit hervorzuheben, im Rahmen derer immer wieder die theoretischen Möglichkeiten im Bereich Gamifizierung, mit dem Arbeitsaufwand der Umsetzung und den in den Praxistests erhobenen Bedürfnissen der Nutzer:innen in Beziehung gesetzt wurden.

Die wichtigsten Ergebnisse im AP 4 sind 1) ein schlüssiges Gamifizierungskonzept für die die Nutzerinteraktion der gesamten Plattform 2) dieses wurde in den die Plattform integriert, 3) ein eigenes Konzept für die Synthese von Beiträgen wurde erstellt und 4) anhand eines interaktiven-Dummies mit Nutzer:innen getestet.

Die wichtigsten Positionen im zahlenmäßigen Nachweis bei LIQD sind die Personalkosten der Entwickler:innen, die die Umsetzung des Prototypens sicherstellten, sowie des Designers und der Projektmanager:innen, die insbesondere in diesem AP Anteil an der Konzeption des Designs hatten sowie den Click-Dummy konzeptionierten, designten und aufsetzten. In der Arbeit des IPG waren die wichtigsten Positionen die Personalkosten für Recherche und Konzeption der Gamifizierungsmechanismen und deren Abstimmung zu LIQD und das Usertesting.

AP 5 Praxistests

Ziel von AP 5 war es, die Prototypen aus AP2 und AP3 mit potenziellen Nutzer:innen zu testen und somit eine praxisnahe Rückmeldung bezüglich der Erreichung der Ziele von KOSMO zu bekommen. Damit konnten zum einen technische Fortschritte der KI und des Dashboards sowie der Funktionsumfang aus Sicht der Nutzer:innen reflektiert und überprüft werden. Die Praxistests stellten ein Element der partizipativen Entwicklung von KOSMO dar und ermöglichten einen kontinuierlichen Einbezug von möglichen Nutzer:innengruppen in die Entwicklung. Sie sollten zudem als Grundlage für die wissenschaftlichen Evaluation des Assistenzsystems dienen.

Im ersten Praxistest (10/2021) lag der Fokus auf den relevanten Vorkenntnissen und Einstellungen der Moderierenden in der Zielgruppe sowie der wahrgenommenen Nutzer:innenfreundlichkeit des KOSMO-Moderationsdashboards und der implementierten KI zur automatischen Erkennung von Inzivilität. Es wurden 12 Testpersonen aus Verwaltung und öffentlich-rechtlichen Nachrichtenmedien mit Vorerfahrung in der Moderation von Online-Plattformen ausgewählt und deren Erlebnis der Plattform zu diesem Zeitpunkt getestet. Diese Personen nahmen an einem User Testing teil, in dem ihnen Aufgaben gegeben wurden, die sie über die Nutzung des Dashboards lösen sollten und somit die Plattform in ihren Funktionen kennenlernten. Sie wurden dabei von einer Person aus dem Forschungsteam begleitet. Anschließend wurde ein Interview zu den Erfahrungen im Umgang mit der Plattform durchgeführt, um die Erfahrung zu reflektieren. Die Konzeption, Durchführung und Auswertung des ersten Praxistests lag federführend bei der HHU. Die Ergebnisse sind im Kapitel “Wissenschaftlicher Fortschritt” (S. 18) dargestellt.

Im zweiten und ebenfalls von der HHU durchgeführten Praxistest wurde der Fokus auf die Wahrnehmung von Diskussionsteilnehmenden in Bezug auf die KI-Kategorien (unangemessen, deliberativ hochwertig, Tatsachenbehauptungen) gelegt. Ausgangspunkt waren die Trainingsdaten für die KI. Ziel war die kritische Prüfung, ob die KI a) die Wahrnehmung potentieller Diskussionsteilnehmender in Bezug auf die KI-Kategorien widerspiegelt und b) sie dies in sozial ausgewogener Weise tut. Die zentrale Fragestellung des Tests war, inwiefern der Bildungshintergrund von Annotator:innen sich auf die Annotation von Beiträgen auswirkt und somit die KI beeinflusst. Dazu wurden im Rahmen der in AP 2 durchgeführten Crowdannotation zusätzlich soziodemografische Daten der Annotator:innen erhoben, um die Biases im Rahmen des Praxistest zu überprüfen. Zusätzlich wurde erstmalig darauf geachtet, dass eine gleichmäßige Verteilung unterschiedlicher formaler Bildungsniveaus der Crowdannotierenden in den erhobenen Daten vorliegt (siehe auch Kapitel “Wissenschaftlicher Fortschritt”, S. 18). Konzeption, Durchführung und Auswertung des Tests lagen vollständig bei der HHU.

In einem abschließenden Praxistest sollte der KOSMO Prototyp zur Moderation einer realen Online-Diskussion durch einen der Praxispartner:innen eingesetzt werden. Ziel war es dabei, die Eignung des Prototyps für die reale Anwendung kritisch zu prüfen sowie belastbare Ergebnisse zur Qualitäts- und Effizienzsteigerung durch die Verwendung von KOSMO zu erhalten. Der dritte Praxistest zielte somit auf ein Erkenntnisinteresse in den Evaluationsdimensionen deliberative Qualität, Moderationseffizienz sowie KI-Akzeptanz von Anwender:innen und Endnutzer:innen ab. Für die Umsetzung des Praxistests führten LIQD und die HHU im Dezember 2022 Gespräche mit dem Bezirk Berlin-Pankow. Der Praxistest kam für den angestrebten Zeitraum im Frühjahr 2023 allerdings nicht zustande (siehe hierfür auch AP7: Akquise von Pilotanwender:innen, S. 15). Dies lag insbesondere an rechtlichen Bedenken zum Einsatz der KOSMO-KI in realen Beteiligungsverfahren, allerdings wurde Interesse signalisiert, den Praxistest mit einem entsprechenden Vorlauf im Rahmen einer langfristigen Kooperation außerhalb des Projektzeitraumes durchzuführen.

AP 6 Evaluation

Ziel des Arbeitspaketes 6 war die wissenschaftliche Evaluation der Entwicklung und des Einsatzes des KOSMO-Assistenzsystems auf Basis der oben beschriebenen Praxistests. Die Evaluation diente dem Zweck, den Beitrag von KOSMO für die wissenschaftliche und praktische Erweiterung des Status quo im Forschungsbereich automatisierter Moderationssysteme und deren Wirkungen auf die Qualität von Online-Diskussionen sichtbar zu machen und Effekte, Potenziale und Grenzen des Prototyps kritisch zu reflektieren. Sie ist demnach essenziell für die Qualitätssicherung im Projekt selbst sowie für die wissenschaftlich-technische Validierung des Prototyps. Die Evaluation erfolgte entlang folgender Dimensionen: (1) Steigerung der deliberativen Qualität von Online-Diskussionen; (2) Steigerung der Effizienz und Wirksamkeit der Moderation von Online-Diskussionen und (3) Steigerung der Akzeptanz und der Usability des Assistenzsystems. Die Evaluation erfolgte sowohl formativ als auch summativ. Die formative Evaluation trägt dem iterativen Arbeitsplan des Vorhabens Rechnung und diente dazu, die Zwischenergebnisse des Projekts in den weiteren Arbeitsplan zu integrieren. Die summative Ergebnisevaluation dagegen hat den Projekterfolg als Ganzes im Blick und ist dazu geeignet, das Erreichen der mit dem Vorhaben verknüpften Ziele zu überprüfen.

Im Rahmen des Arbeitspaketes 6 wurden folgende Leistungen erbracht: Zunächst wurde ein Evaluationskonzept angefertigt, welches die oben genannten Evaluationsdimensionen konkretisiert, operationalisiert und in Studiendesigns überführt hat. Die weiteren Arbeiten im AP konzentrierten sich auf die Konzeption, Durchführung und Auswertung der vorgesehenen Praxistests. Diese beinhalteten die User-Testings (Praxistest 1), die im Oktober 2021 durchgeführt wurden. Anschließend wurde ab Juni 2022 mit der Konzeption, Durchführung und Auswertung der Begleitstudie zur Crowdannotation (Praxistest 2) in Kooperation mit AP 2 begonnen. Für den 3. Praxistest wurde ein wissenschaftliches Studiendesign für die Evaluation des Moderationsdashboards in der Praxis vorgelegt, welches im Rahmen eines wissenschaftlichen Folgeprojekts außerhalb des Projektzeitraums genutzt werden kann. Informationen zu Konzeption und Durchführung der Praxistests finden sich im vorangegangenen Abschnitt zum AP 5. Abschließend wurde der Evaluationsbericht (summative Evaluation) angefertigt und im Rahmen eines Abschlussworkshops im März 2023 präsentiert.

Die Endevaluation machte deutlich, dass der größte Erkenntniszuwachs in der Dimension “Technologieakzeptanz und Usability” erzielt werden konnte. Der Forschungsstand zu Online-Deliberation und -Moderation konnte hier durch die Anknüpfung an Erkenntnisse in den Bereichen 1) Technologie- und KI-Akzeptanz sowie 2) KI-Nutzung im öffentlichen Sektor substanziell erweitert werden. So wurde durch die Erhebung von Fähigkeiten und Voreinstellungen für den Einsatz von KI-gestützter Moderation von Moderator:innen im öffentlichen Sektor im Rahmen des ersten Praxistests eine bislang vernachlässigte Zielgruppe erschlossen. Die Einblicke tragen dazu bei, dass sowohl Produktentwicklung als auch Implementierung von KI im öffentlichen Sektor zukünftig bedarfsgerechter und erfolgsversprechender durchgeführt werden können. Die Einblicke in die Vorstellungen zu Transparenz von KI-Entscheidungen als zentrales Merkmal zur Akzeptanzsteigerung des Einsatzes KI-gestützter Moderation bieten außerdem Anknüpfungspunkte an das Forschungsfeld der Explainable AI. Der nennenswerteste Befund auf der Dimension der “deliberativen Qualität” wurde durch den zweiten Praxistest erbracht. Durch die Crowdannotation wurde die Wahrnehmung von unangemessenen und hochwertigen Kommentaren potenzieller Diskussionsteilnehmende umfänglich und sozial inklusiv in das Training der KI integriert. Als größte Limitation muss festgehalten werden, dass durch den Wegfall des dritten Praxistests keine Aussagen darüber getroffen werden können, inwiefern der Einsatz von KOSMO die deliberative Qualität von Diskussionen beeinflusst oder zur Verbesserung der Moderationseffizienz

beiträgt. Eine Evaluation außerhalb des Projektzeitraumes wird aber durch die Schnittmengen mit dem ebenfalls an der HHU angesiedelten Drittmittelprojekt (Manchot-Forschungsgruppe) “Entscheidungsfindung mit Hilfe von Methoden der Künstlichen Intelligenz” im Rahmen des sozialwissenschaftlichen Teilprojekts “Unterstützung politischer Entscheidungen durch Künstliche Intelligenz“ (UPEKI) umgesetzt (siehe hierzu auch Teil 3, S. 23). Eine detaillierte Beschreibung der wissenschaftlichen Ergebnisse des Projektes findet sich im Teil zum wissenschaftlichen Fortschritt im Rahmen des Projektes sowie im wissenschaftlichen Abschlussbericht des AP6. Über den gesamten Projektzeitraum hinweg wurden die erzielten Ergebnisse nicht nur dem Konsortium, sondern auch der wissenschaftlichen Fachöffentlichkeit durch die Erarbeitung von Präsentationen und Publikationen zugänglich gemacht (siehe AP7, S. 12ff.).

Insgesamt konnten fast alle Arbeitsschritte und Meilensteine im AP mit den oben skizzierten Änderungen erreicht werden. Dazu gehört die 1) Entwicklung und dynamische Anpassung eines Evaluationskonzepts; 2) die Vorbereitung uns Auswertung der Anforderungsanalyse (siehe AP 1); 3) Evaluation und Dokumentation der Praxistests sowie 4) die Erstellung eines Evaluationsberichts, der die zentralen Ergebnisse des Projekts zusammenfasst und Handlungsempfehlungen formuliert.

Als wichtigste Position des zahlenmäßigen Nachweises in AP6 ist die Beschäftigung der wissenschaftlichen Mitarbeiterin zu nennen. Diese Ausgaben waren notwendig, um die im Arbeitspaket anfallenden Aufgaben zu erfüllen. Da durch die wissenschaftlichen Zwischenevaluationen Potential für Synergien zwischen den Arbeitspaketen zutage traten, hat das AP6 an verschiedenen Stellen die Arbeit anderer Arbeitspakete zusätzlich substanziell unterstützt (insbesondere AP5, siehe hierfür Anmerkungen in diesem Teil des Abschlussberichts). Die Erledigung dieser und aller zusätzlicher Aufgaben hat substanziell zum Fortkommen des Projektes beigetragen. Im Zuge dieser Tätigkeiten fielen zudem kleinere Positionen an, wie z.B. die Einstellung einer studentischen Hilfskraft zur Unterstützung der Auswertung der Auftaktworkshops und des User-Testings und die Anschaffung einer Transkriptionssoftware.

AP 7 Ergebnisverbreitung und Verbreitungsmodelle

Ziel des Arbeitspaketes war die Bekanntmachung von KOSMO und die Konzeption tragfähiger Verbreitungsmodelle für eine weiter fortgeschrittene Version. Dazu wurden unterschiedliche Arbeitsschritte geplant: (1) Projektpräsentation, (2) Publikation und Fachveröffentlichungen, (3) Analyse gängiger und innovativer Verbreitungsmodelle, (4) Entwicklung möglicher Verbreitungsmodelle, (5) Akquise von Pilotanwender:innen. Neben den wissenschaftlichen Publikationen zu Teilergebnissen der Forschung an KOSMO (siehe auch Punkt “Erfolgte und geplante Veröffentlichungen”, S. 21) wurden für die Präsentation des Projektes in der Öffentlichkeit Informationsmaterialien gestaltet, die über das Vorhaben als Ganzes und die zu Grunde liegende Vision informieren. Diese Materialien sind bewusst niedrigschwellig und einfach gehalten. Dies waren konkret: Eine Website, eine Drucksache, ein mehrstündiges Event und ein Erklär-Video.

Projektpräsentation & Ergebnisverbreitung

Mit der Website sollen alle interessierten Personen die Möglichkeit haben, sich niedrigschwellig über KOSMO zu informieren. Dazu werden dort die Grundlagen von KOSMO mit bildlich-illustrativen Mitteln vorgestellt. Die Website erklärt, inwiefern künstliche Intelligenz und ein nutzerzentriert gestaltetes Backend die Moderation von Diskussionen im digitalen Raum unterstützen können. Dieser Überblick wird inhaltlich mit Texten zu unseren Forschungsbausteinen Workshops & Partizipative Erarbeitung, KOSMO & adhocracy, Qualitätskriterien guter Diskussionen im Netz und Gamification ergänzt. Die Website bleibt über das Projekt hinaus als Kommunikationskanal bestehen und wird zudem genutzt,

um die Ergebnisse des Projektes zu veröffentlichen (Open Access). Dies betrifft sowohl diesen Bericht als auch – soweit rechtlich möglich - die weiteren unten aufgeführten Publikationen. Für weitere Anfragen, Austausch und Feedback können die Konsortialpartner über die Website kontaktiert werden. Die Inhalte der Website sind ebenfalls in einer Drucksache analog verfügbar. Damit können auch auf Tagungen und bei Vor-Ort-Terminen Informationen zu KOSMO weitergegeben werden. In kurzen Videos stellen wir in einfachen Worten die Betaversion und die Frontend- und Backend-Funktionen von KOSMO vor. Ebenfalls wird mit Hilfe der Videos erklärt, wie die KI die Moderation auf der Plattform unterstützt.

Am 15.11.2022 wurde KOSMO öffentlich in einer Veranstaltung vorgestellt und mit interessierten Gäst:innen und Praxispartner:innen getestet. Dafür wurden seitens des Projektteams umfangreiche Testumgebungen bereitgestellt, die die gesamte Funktionalität demonstrierten. In einer anschließenden Diskussion wurde mit den Teilnehmenden die Frage erörtert, inwieweit KI einen unterstützenden Beitrag in demokratischen Prozessen leisten kann.

Die Mitarbeitenden der HHU präsentierten die Forschungsergebnisse auf internationalen wissenschaftlichen Fachtagungen und fertigten im Projektzeitraum mehrere publikationsreife Studien an (S. 21). So wurden die Projektergebnisse mit der wissenschaftlichen Community diskutiert und sichtbar gemacht.

Analyse gängiger und innovativer Verbreitungsmodelle

Die Analyse gängiger und innovativer Verbreitungsmodelle wurde zunächst mittels einer Marktanalyse bei Anbieter:innen vergleichbarer Softwareangebote begonnen. Dabei erwiesen sich folgende Faktoren als besonders relevant: 1) die Wege, auf denen das Produkt angeboten wird, 2) Preismodelle, 3) Das Vertriebsmodell (z.B. kostenlose Erstberatung, Demoversion etc.), 4) Möglichkeiten der Anwendung (Integration in eigene Software möglich, Nutzung nur möglich gepaart mit honorierter Begleitung durch Anbieter:in). Die Analyse wurde durch Literaturrecherche untermauert. In der Marktanalyse und der Literatur wurde ein starker Trend zu SaaS-Angeboten (Software as a Service) festgestellt.

Im Kontext von KI-Anwendungen wurde schnell deutlich, dass wirtschaftlich tragfähige Verbreitungsmodelle für KI-Anwendungen in der Breite momentan noch in der Entwicklung sind und sich am Markt noch etablieren müssen. Auf dem Gebiet der Software für Bürger:innenbeteiligung sind diese bisher noch überhaupt nicht zu finden. KOSMO nimmt hier eine Pionierstellung ein.

Um die Recherche zu untermauern, wurden Interviews mit Praxispartner:innen durchgeführt, um deren use cases und Ansprüche an ein Verbreitungsmodell für KOSMO herauszuarbeiten. Wesentliches Ergebnis war, dass durch die Nutzung eigener Plattformen vor allem eine Integration von KOSMO über eine bereitgestellte API interessant ist. Gleichzeitig ist gerade in der Kooperation mit kommunalen Anwender:innen das Thema Datenschutz besonders wichtig und mit besonderer Intensität zu behandeln. Hier sind viele Einzelfälle zu unterscheiden. Dafür ist vor einem Einsatz eine breite Anpassung der Datenschutzregelungen auf den Plattformen notwendig und von den Nutzer:innen abzeichnen zu lassen. Dies erhöht die Hürden für die Verbreitung von KOSMO deutlich.

In einem internen Workshop (IPG/LIQD) wurden aufbauend auf den Analysen über ein strukturiertes Value Proposition Design verschiedene Zielgruppen und deren Nutzungsanforderungen an KOSMO, und anschließend pains (Herausforderungen) und gains (wie kann KOSMO diesen entgegnen) herausgearbeitet. Abschließend wurde ein wünschenswertes KOSMO-Produktversprechen formuliert und überprüft, inwieweit dieses bereits jetzt erfüllt werden kann, und wo die Entwicklung der Software

dafür fortgeführt werden sollte. Im Ergebnis wurden die wesentlichen notwendigen Bausteine für ein wirtschaftlich verwertbares Produktversprechen von KOSMO erarbeitet.

Entwicklung Verbreitungsmodelle

Auf Basis des Produktversprechens und der vorangegangen Analyse wurden mittels Business Model Canvas zwei mögliche Verbreitungsmodelle entwickelt. Mittels SWOT-Analyse wurden diese auf Stärken und Schwächen hin überprüft. Schlüsselergebnis war hier: Der wesentlich limitierende Aspekt, der in beiden Modellen zum Tragen kommt und eine hohe Hürde für eine nachhaltige und langfristige Verbreitung darstellt, sind die hohen Kosten für die Annotation der Trainingsdaten und die Weiterentwicklung der KI. Insbesondere der Ansatz, eine nach hohen gemeinwohlorientierten und wissenschaftlichen Standards programmierte KI zu entwickeln und die Einhaltung von hohen ethischen Standards auch in der Schulung und Entlohnung derjenigen, die die Trainingsdaten aufbereiten, bilden eine schwer zu überwindende Hürde. Darüber hinaus sind anhand des vorliegenden Prototyps die Kosten für den Produktiv-Betrieb der KI schwer vorauszusehen. Dies führt zu weiteren Unsicherheiten in der Entwicklung tragfähiger Verbreitungsmodelle.

Dem steht auf Nutzer:innenseite der mit KOSMO angedachte Anwendung bisher kein ausreichender ROI (Return on Invest) gegenüber, der die aus der Softwareentwicklung und Annotation von Trainingsdaten entstehenden hohen Kosten tragen könnte. Im Folgenden erläutern wir die Schwerpunkte der beiden Verbreitungsmodelle:

1) KOSMO als Integration in adhocrcay+: Die KI-Funktionalität von KOSMO wird als Teil der Plattform mit verbreitet und erweitert deren Einsatzspektrum. Vorteil: Klares Produkt mit einer klaren Zielgruppe und Liquid Democracy als Vertrieb, sowie kontinuierliche Softwareupdates über adhocracy+ und ein bestehendes Supportangebot. Nachteil: Keine Integration in andere Systeme möglich und dadurch eine eingeschränkte Nutzer:innengruppe. Dies führt neben den begrenzten wirtschaftlichen Skaleneffekten (ausschließlich Kund:innen von Beteiligungssoftware als Zielgruppe) zusätzlich zu einer eingeschränkten Datengrundlage für die weitere KI-Entwicklung.
2) KOSMO als API: Die KI-Funktionalität von KOSMO und wenn möglich auch das innovative Moderations-Dashboard werden per API (Application Programming Interface) an bestehende Systeme angebunden, die Online-Debatten ermöglichen. Vorteile: Kann von einer großen Zielgruppe genutzt werden, da die API an bereits bestehende Plattformen angeschlossen werden kann und individuelle Anpassungen möglich sind. Diese Herangehensweise bietet gute Skalierungschancen und dadurch eine große Basis potentieller zahlender Nutzer:innen. Nachteile: Kosten für individuelle Anpassungen auf Seiten der Anwendersoftware für die potentiellen Nutzer:innen. Dies kann nicht von Liquid Democracy betreut werden. Aus ethischer Sicht besteht das Risiko, dass KOSMO für Zwecke genutzt wird, die nicht den Werten des Projekts entsprechen. Es ergibt sich ein höherer Aufwand zum Training der KI als in Verbreitungsmodell 1, da die Datengrundlage für eine allgemeine Verwendung deutlich größer werden müsste. Damit wirtschaftlich noch schwerer umzusetzen als Verbreitungsmodell 1.

Aus der Arbeit aus diesem Arbeitspaket lässt sich die grundlegende weitere Forschungsfrage ableiten, wie eine gemeinwohlorientierte Open-Source-KI wirtschaftlich tragfähig zu entwickeln und vertreiben ist. Durch die komplexen Zusammenhänge zwischen wissenschaftlicher Fundierung, ethischen und gemeinwohlorientierten Standards, Datenschutz und der notwendigen wirtschaftlichen Basis bietet sich diese als Ansatzpunkt für ein weiterführendes Forschungsvorhaben an.

Akquise Pilotanwender:innnen

Das Interesse an der Anwendung von KOSMO war seitens verschiedener Kommunen und Praxispartner hoch. So wurde beispielsweise im Herbst 2022 das Bezirksamt Pankow als potenzielle Pilotanwender:in akquiriert. Die dazu geführten Gespräche und Präsentationen verliefen vielversprechend, erbrachten aber kein Ergebnis. So konnten schlussendlich im Projektzeitraum keine Pilotanwendungen durchgeführt werden. Wir haben folgende Erkenntnisse, warum dieser Punkt nicht erfolgreich war:

1. Aufgrund von Bindungen an bereits verwendete Plattformen zur Bürgerbeteiligung sowie Migrations- und Datenschutzproblematiken, die im Projektzeitraum nicht gelöst werden konnten, konnten diese Kommunen/Praxispartner nicht auf die KOSMO-Plattform wechseln.
2. Eine „Test-Beteiligung" im Sinne einer „Sandbox" wiederum wurde von Teilnehmenden aufgrund der fehlenden Relevanz nicht angenommen. Das Testen von Online-Beteiligung anhand eines hypothetischen Beteiligungsgegenstandes hat eine sehr geringe Attraktivität für Teilnehmende aus der Bürgerschaft, so dass die Ergebnisse nicht die für eine Auswertung notwendige Reife erzielt hätten.
3. Eine einzelne Beteiligung im Rahmen eines Prototyps außerhalb der eigentlichen Beteiligungsplattform einer Kommune bzw. eines Praxispartners aufzusetzen, stellt diese vor massive Kapazitätsprobleme in der Betreuung. Die dafür notwendigen administrativen Arbeiten können kaum mit dem meist ausgelasteten Personal bewältigt werden. Dies wäre anders, wenn die Software nach dem Test direkt in den dauerhaften Betrieb übernommen würde. Dies schließt die Bindung (siehe 1.) an andere Plattformen und der Entwicklungszustand von KOSMO auch beim Verbreitungsmodell (s.o.) jedoch aktuell aus.

Die praktische Anwendung von KOSMO im Rahmen eines großen drittmittelfinanzierten Feldexperiments (UPEKI, siehe auch AP 6), konnte in der Projektlaufzeit jedoch vorbereitet werden und wird im weiteren Verlauf des Projektes durch die HHU begleitet. Die Erkenntnisse aus diesem Arbeitspaket bieten wertvolle Einsichten für Folgeprojekte: So ist die Anwendung von Prototypen in diesem Bereich eher im (vor-)wissenschaftlichen als im realweltlichen Kontext umsetzbar, was in Folgeprojekten entsprechend eingeplant werden sollte.

In AP 7 sind die wesentlichen Aufwände des IPG die 1) Konzeption und Gestaltung der Materialien zur Projektpräsentation, 2) die Analysen, Konzeptionen und Recherchen für Verbreitungsmodelle, 3) die Begleitung und Durchführung des öffentlichen Events und verschiedener Workshops und Konsortialtreffen. Die Ausgaben für die AP7 vonseiten der HHU umfassten die Übernahme von Reisekosten zur Präsentation der Ergebnisse auf nationalen und internationalen Fachtagungen (insbesondere im Rahmen der 9th European Communication Conference (ECREA) 2022 in Aarhus) sowie zur Teilnahme an internen Workshops in Präsenz und dem Launch-Event. Für Liquid Democracy sind die wesentlichen Positionen des zahlenmäßigen Nachweises für das AP 7 Personalkosten zur Teilnahme an internen Workshops im Rahmen des AP 7, sowie Reise- und Personalkosten für die Teilnahme an Veranstaltungen zur Vorstellung von KOSMO.

Wesentliche Ergebnisse dieses Arbeitspaketes umfassen 1) die Darstellung des Projektes in niedrigschwellig erfassbaren Medien, 2) die Präsentation des Projektes zu verschiedenen Gelegenheiten (s.u.), 3) die Recherche, Erfassung und Analyse gängiger Verbreitungsmodelle, 4) die Konzeption eigener Verbreitungsmodelle und die Erfassung ihrer Stärken und Schwächen.

AP 8 Projektmanagement und -koordination

Das AP 8 hatte zum Ziel, einen reibungslosen Ablauf des Projektes innerhalb des Konsortiums zu sichern, damit die Ziele des Vorhabens eingehalten werden. Das AP 8 hatte dazu zur Aufgabe, die Vorhabenskoordination und die Qualitätskontrolle zu gewährleisten. Im Vorhaben waren außerdem halbjährliche Berichterstattungen an den Träger vorgesehen, auf die aber nach Rücksprache mit dem Träger einvernehmlich verzichtet wurde. Die Organisation, Moderation und Durchführung des AP 8 lag ausschließlich bei LIQD.

Das AP 8 war maßgeblich für die gemeinsame und co-kreative Arbeitsweise im Konsortium verantwortlich. Über zweiwöchentliche Jour Fixe sowie digitale Projektmanagement- und Kommunikationstools wurde regelmäßig ein Update über den Stand der jeweiligen Arbeitspakete gegeben und sich über Fortschritt und Hindernisse ausgetauscht. Jeweils zum Ende einer Iteration wurde der Status des Projektes analysiert, mögliche Risiken identifiziert und daraufhin der Zeitplan angepasst. Das ermöglichte dem Konsortium, schnell auf Risiken und Herausforderungen reagieren zu können und auch angesichts von Hindernissen stetig das Projekt weiterzuentwickeln. Bei Bedarf fand eine Anpassung des Zeitplans auch innerhalb der Iteration statt.

Der Ansatz des Projektmanagements, der sich auf die Projektmanagement-Methode Scrum bezieht, hatte außerdem den Effekt, dass alle Konsortialpartner:innen über den Stand des Vorhabens sowie über anstehende Herausforderungen informiert waren. Die Pflege und das Monitoring des dynamischen Projektplans sowie des Dokumentensharing-Systems waren dabei wichtige Aspekte zur Sicherstellung der Dokumentation. Außerdem wurden regelmäßig die Perspektiven des gesamten Konsortiums oder je nach Aufgabe zwischen zwei Projektpartnern eingeladen und beachtet. So herrschte eine Atmosphäre einer co-kreativen Arbeitsweise, in der eine gute Balance aus Fachexpertise und Blick von „außen“ entstand. Die Konsortialpartner waren sich dabei kollegiale Berater:innen, die sich unterstützten, die Fachexpertise gemeinsam in das große Ganze einzuordnen. Orientierung gab dem Konsortium dabei die gemeinschaftliche Erarbeitung einer Vision und eines Wertegerüstes für das Projekt auf Basis der Vorhabenbeschreibung.

Wesentliche Posten im zahlenmäßigen Nachweis bei Liquid Democracy sind die Personalkosten für die Konsortialführung, die Projektleitung und das Projektmanagement im Rahmen der Organisation des Gesamtprojektes. Dabei wurden folgende geplante Ergebnisse umgesetzt: (1) Dynamischer Projektplan und Arbeitspakete Status Management System, (2) Projektbericht und Zusammenstellung der Ergebnisse der Arbeitspakete. Die halbjährliche Berichterstattung in Form von Konsortialberichten wurde in Absprache mit dem Träger nach einer einmaligen Einreichung eingestellt.

3. Nutzen und Verwertbarkeit für die Zukunft

Zum Projektende besteht, wie geplant, der Prototyp eines KI-gestützten Assistenzsystems für Online-Diskussionen auf Basis der Open-Source-Software Adhocracy, der die Effizienz und Effektivität der Moderation von Online-Diskussionen und -Partizipationsverfahren steigern kann. Die im Prototyp integrierten KI-Algorithmen zur Detektion von qualitativ minder- und hochwertigen Diskussionsbeiträgen sind auch nach Abschluss des Projekts unseres Wissens für Deutschland einzigartig differenziert, transparent und wissenschaftlich fundiert. Wie sich in der Kalkulation der Verbreitungsmodelle herausstellte, liegt eine Herausforderung in der nachhaltigen Finanzierung des derzeitigen Prototypens und seiner Fortentwicklung. Wirtschaftlich tragfähige Verbreitungsmodelle für KI-Anwendungen sind in der Breite momentan noch in der Entwicklung und müssen sich am Markt noch etablieren, während sie

auf dem Gebiet der Beteiligungssoftware bisher gar nicht zu finden sind. KOSMO nimmt hier eine Pionierstellung ein. Ziel ist es deshalb, mithilfe einer Anschlussförderung von KOSMO die Weiterentwicklung voranzutreiben und eine nachhaltige Finanzierung eines Assistenzsystems mit einer ethischen und nachhaltig gepflegten KI zu sichern. Basis dafür sind insbesondere die Fortschritte im technischen und wissenschaftlichen Gebiet des Vorhabens.

Technischer und wissenschaftlicher Fortschritt auf dem Gebiet des Vorhabens

Zum Ende unseres Projektes haben auf dem Gebiet der KI- Entwicklung starke Veränderungen stattgefunden, beispielsweise durch den Release von ChatGPT4 im März 2023. Bei diesen Weiterentwicklungen stellt sich allerdings die Frage, inwieweit diese Standards zu Datensparsamkeit und Orientierung am demokratischen Wertesystem erfüllen. Die Entwicklung von KOSMO unterbreitet hierzu einen Gegenvorschlag, der die KI Anwendung entlang demokratischer Standards in deutschsprachigen Online-Öffentlichkeiten ermöglichen soll. Technisch sowie wissenschaftlich hat das Vorhaben KOSMO im Bereich der Moderation von Online-Diskussionen diesbezüglich starke Fortschritte gemacht. Im Folgenden werden die Erfolge konkreter aufgeschlüsselt.

Technischer Fortschritt

Aus technischer Perspektive erzielt das Projekt substanziellen Fortschritte hinsichtlich der folgenden drei Bedarfe: 1) Der Entwicklung von Künstlicher Intelligenz für die Moderation von Online-Diskussionen. 2) Der Verfügbarmachung von für die Forschung zugänglichen Ressourcen und Instrumenten für die automatische Erkennung von Inzivilität und Deliberation für deutschsprachige Inhalte. 3) Der Erweiterung von Softwarelösungen um KI-Komponenten zur Unterstützung der manuellen Moderation von Online-Diskussionen. Hierbei hat sich der trans- und interdisziplinäre Forschungsansatz an der Schnittstelle zwischen Sozialwissenschaften, Computer Science und Akteuren der Zivilgesellschaft als besonders sinnvoll für die Entwicklung und Evaluation der Machine-Learning-Algorithmen herausgestellt, da dieser nicht nur die technische Optimierung, sondern auch den sinnvollen und zielgerichteten Einsatz von KI-basierten Systemen im Blick behält. Folgende übergeordnete Achievements konnten im Rahmen des Forschungsprojekts in diesem Sinne erzielt werden:

> ***Achievement 1:*** *Die aus der Demokratietheorie abgeleiteten Konzepte Inzivilität und Deliberation wurden für die Erkennung durch Machine-Learning-Algorithmen in einem Praxissetting operationalisiert und in Trainingsdaten übertragen.*

Inzivilität und Deliberation sind auf Demokratietheorien basierende Konzepte, die sich für die Beschreibung und Analyse von “schädlichen” und “wertvollen” Beiträgen in Online-Diskussionen eignen, jedoch nicht ohne Weiteres in den Praxisgebrauch übertragbar sind. Zum Beispiel haben Nutzende und Moderator:innen oft ein intuitiveres Verständnis von Qualität und die Verletzung von Diskussionsnormen und -regeln wird durch plattformspezifische Nettiquetten teilweise unterschiedlich gehandhabt. KOSMO ist es für den deutschsprachigen Raum erstmals gelungen, bislang vor allem in wissenschaftlichen Forschungskontexten verwendete Qualitätskonzpte (Deliberation, Inzivilität) in ein Praxissetting zu überführen und damit potenziell fundiertere und intersubjektiv nachvollziehbarere Moderationsentscheidungen zu fördern.

> ***Achievement 2:*** *Optimierung von Machine-Learning-Algorithmen für die Erkennung von Diskussionsbeiträgen in deutschsprachigen Online-Diskussionen.*

Von der kontinuierlichen Entwicklung immer leistungsstärkerer KI-Modelle und Algorithmen profitiert auch die KI-basierte Moderation. Jedoch werden neue Algorithmen zunächst meist für den englischen Sprachraum entwickelt und sind somit für deutschsprachige Diskussionen nicht anwendbar. Oft müssen geeignete Modelle neu für jeden Anwendungsbereich entwickelt oder zumindest aufwendig angepasst werden. Zudem unterscheidet sich die Eignung verschiedener KI-Modelle je nach Praxisanwendung. Beispielweise muss die Leistungsfähigkeit oft mit Fragen nach Ressourcen und Transparenz abgewogen werden. Im Rahmen von KOSMO wurden moderne KI-Algorithmen aufwendig für den Use Case Online-Moderation im deutschen Sprachraum weiterentwickelt und für die wissenschaftliche Fachcommunity verfügbar gemacht.

> ***Achievement 3:*** *Implementierung von KI in Moderationstool als Leuchtturmbeispiel für positiv-demokratischen und ethischen Einsatz.*

Derzeitige KI-Systeme werden hauptsächlich in ökonomisierten Kontexten eingesetzt, z.B. zur Akquirierung von Konsument:innen. Ziel von KOSMO war es, KI zur Unterstützung besserer Diskurse in demokratischen (Online-)Räumen einzusetzen. KOSMO schafft es, KI nicht als Bedrohung, sondern als aktive Unterstützung des demokratischen Wertesystems zu verstehen. Diese Resonanz bekamen wir regelmäßig in Projekt- und Prototyps-Präsentationen sowie in Praxistests. Bestärkt wurde dies durch das datensparsame Training der KI, in dem keine sensiblen Daten verarbeitet wurden. Diese positive Umdeutung von technischem Fortschritt als Möglichkeit zu demokratischer Weiterentwicklung ist aus demokratiebildender Perspektive von großer Wichtigkeit und leistet einen Beitrag zum gesellschaftlichen Diskurs einer fairen und nachhaltigen KI in demokratischen Prozessen.

Wissenschaftlicher Fortschritt:

Aus wissenschaftlicher Perspektive liegt der zentrale Ertrag des KOSMO-Projekts in der Erlangung fundierter empirischer und angewandter Erkenntnisse im Hinblick auf den Einsatz KI-gestützter Moderationssysteme in Online-Diskursen. Dies wurde durch substanzielle Erkenntnisgewinne sowohl entlang einer technischen als auch sozialwissenschaftlich-theoretischen Dimension geleistet. Aus technischer Sicht wurden erstmalig Machine-Learning-Algorithmen zur automatisierten und differenzierten Klassifikation von inzivilen Diskussionsbeiträgen (u.a. beleidigende, zynische und hasserfüllte Beiträge) und deliberativ hochwertige Beiträge (u.a. argumentative, tatsachenbeanspruchende und lösungsorientierte Beiträge) in deutschsprachigen Online-Diskussionen entwickelt und in ein Moderationsdashboard implementiert. Aus sozialwissenschaftlicher Perspektive wurde der Forschungsstand zu Online-Deliberation und Moderation weiterentwickelt, indem er an Forschungserkenntnisse zu 1) Technologie- und KI-Akzeptanz sowie 2) KI-Nutzung im öffentlichen Sektor anschlussfähig gemacht wurde. Im Rahmen einer Bestandsaufnahme zu KI-Vorkenntnissen von Moderator:innen sowie deren Anforderungen an KI-gestützte Moderationssysteme im Kontext von Öffentlichkeitsbeteiligung durch Verwaltung und öffentlich-rechtlichen Medien konnte sichergestellt werden, dass das entwickelte Moderationssystem den Arbeitsanforderungen sowie den hohen demokratischen Standards in diesem Feld genügt. Die geleistete Arbeit trägt damit dazu bei, dass sowohl die Produktentwicklung als auch die Implementierung von KI im öffentlichen Sektor zukünftig bedarfsgerechter und erfolgsversprechender ablaufen können.

Im Rahmen der Entwicklung des KOSMO-Prototyps konnten im Detail folgende Erkenntnisse erzielt und Leistungen erbracht werden:

> ***Achievement 1:*** *Moderator:innen unterstützen mit überragender Mehrheit den Einsatz eines KI-gestützten Assistenzsystems zur Selektion moderationswürdiger Kommentare in der vorliegenden Form.*

Im Rahmen des Auftaktworkshops konnte auf Grundlage der Auswertung der HHU festgestellt werden, dass ein transparenter und fairer Einsatz von KI zur Moderation von Online-Diskussionen grundsätzlich von allen befragten Praxispartner:innen begrüßt und gewünscht wurde. Somit ist der Bedarf und die Nützlichkeitseinschätzung von KI-gestützten Moderationstools grundsätzlich gegeben. Allerdings ist dies an die Voraussetzungen geknüpft, dass die KI (a) in bestehende Arbeitsprozesse eingebunden werden kann, (b) zur Strukturierung und Aufbereitung der Kommentare für Moderierende beiträgt sowie (c) Vorschläge zur Moderation unterbreitet. Eigenständige Moderationsentscheidungen sollen nach Einschätzung der Praxispartner ausdrücklich nicht durch die KI getroffen werden. Die Nützlichkeitseinschätzung der in KOSMO entwickelten Moderations-KI wurde im nachfolgenden Praxistest nochmals durch die testenden Moderator:innen aus der Zielgruppe des öffentlich-rechtlichen Rundfunks und der öffentlichen Verwaltung bestätigt.

> ***Achievement 2:*** *Die angebotenen Machine-Learning-Algorithmen zur Erkennung von unangemessener und hochwertiger Kommunikation entsprechen den Bedarfen von Moderator:innen.*

Ergänzend wurde das Verständnis eines qualitativ hochwertigen Diskurses der Anbieter:innen von Online-Diskussionen sowie deren Anforderungen an KI-gestützte Moderationstools im Rahmen der Auftaktworkshops umfänglich erhoben. Die Erkenntnisse waren forschungsleitend für die Entwicklung der differenzierten KI-Kategorien im Rahmen des AP2. So konnte sichergestellt werden, dass die durch die KI klassifizierten Kommunikationsinhalte auch tatsächlich mit dem Verständnis qualitativ hoch- und minderwertiger Diskurse der Moderator:innen übereinstimmen und so den Anforderungen der Anwender:innen im Arbeitskontext entsprechen.

> ***Achievement 3:*** *Das Moderationsdashboard wird von der Zielgruppe der Moderator:innen als übersichtlich, konsistent und intuitiv bewertet.*

Durch ein umfassendes User-Testing zwischen dem ersten und zweiten Iterationsschritt konnte sichergestellt werden, dass der Prototyp des Assistenzsystems den Usability-Anforderungen der Moderator:innen an die Benutzerfreundlichkeit entspricht, das heißt, dass das System geeignet ist, typische Aufgaben im Anwendungskontext mit Effektivität, Effizienz und zu hoher Zufriedenheit erledigen zu können. Die Success Scores des Enhanced Cognitive Walkthroughs weisen darauf hin, dass Moderator:innen Aufgaben rund um Orientierung auf dem Dashboard, Kommentarprüfung und Kommentarmoderation mithilfe des Moderationsdashboards erfolgreich erledigen konnten. Den anschließenden qualitativen Interviews konnte entnommen werden, dass Moderator:innen das Moderationsdashboard überwiegend als übersichtlich, konsistent und intuitiv bewerteten.

> ***Achievement 4:*** *Es gibt Hinweise darauf, dass durch das Angebot zielgruppendifferenzierter Features zur Schaffung von Transparenz für Moderator:innen a) die Akzeptanz KI-gestützter Assistenzsysteme sowie b) mittelbar die Moderationseffizienz gesteigert werden kann.*

Öffentliche Verwaltung und öffentlich-rechtliche Nachrichtensender als Hauptzielgruppen von KOSMO zeichnen sich durch hohe demokratische Ansprüche an ihre Moderationspraktik aus. Dies äußert sich sowohl in der Betonung rechtlicher Verpflichtungen zur Löschung strafbarer Inhalte als auch der normativen Anforderung eines freien und gleichberechtigten Diskurses für alle Teilnehmenden. In dem Kontext betonen alle Teilnehmende des User Testings, dass die Transparenz der KI-Anwendung ein entscheidender Faktor für die erfolgreiche Anwendung im Arbeitskontext ist. Auf Basis der Daten des

User-Testings konnten erstmals spezifische Vorstellungen von Transparenz für die Zielgruppe der Moderation im Rahmen von (demokratischer) Online-Moderation erhoben werden. Hierzu wurde zunächst ermittelt, welchen Zweck eine transparentes KI-gestütztes System im Arbeitskontext erfüllen sollte. Es konnten folgende Zwecke ermittelt werden: (1) Begründung von KI-Entscheidung im individuellen Anwendungsfall, (2) Entscheidungshilfe zur schnelleren Moderation, (3) Kontrolle der KI, (4) Kommunikation mit moderierten Nutzer:innen und (5) Kommunikation mit der Community. In einem nächsten Schritt wurden dann zwei verschiedene Transparenzanwendungen als Mock-up präsentiert und den Moderator:innen zur Bewertung vorgelegt. Hier zeigte sich, dass die Mock-ups – abhängig von Moderator:innenmerkmalen – als unterschiedlich effektiv bewertet wurden. Nutzende mit geringem bis durchschnittlichem Vorwissen zum Thema KI, die Ihre Rolle primär in der Moderation von Kommentaren begriffen haben, präferierten die Transparenzanwendung mit Kennzeichnung ausgewählter Textstellen. Auf diesem Wege konnten besonders ausschlaggebende Worte für die KI-Klassifikation durch eine optische Hervorhebung identifiziert werden, was sowohl die Begründung von Entscheidungen im Einzelfall, die Beschleunigung der Moderation sowie die Kommunikation mit moderierten Nutzenden erleichterte. Moderator:innen, die ein hohes Fachwissen zum Thema KI und eine hohe Technikbereitschaft vorweisen konnten sowie über Moderationsaufgaben hinausgehende berufliche Aufgaben im beruflichen Kontext innehatten, sprachen sich für die Verwendung von solchen Hervorhebungen sowie zusätzlich einer Feedback Funktion als Transparenzfunktion aus. Dies führe dazu, dass Moderator:innen die KI mithilfe des eigenen Inputs verbessern konnten und so aus Sicht der Anwender eine gewisse Kontrolle über die KI hergestellt werden konnte. Die Befunde weisen darauf hin, dass eine zielgruppendifferenzierte Herangehensweise an die Schaffung von Transparenz von KI effektiver zur Erfüllung der dahinterliegenden Bedarfe beitragen könnte und dass Voraussetzungen und Berufsverständnis der Moderator:innen eine entscheidende Rolle dabei spielen. Diese Erkenntnisse bereichern den wissenschaftlichen Diskurs um die KI-Akzeptanz verschiedener Personengruppen signifikant.

> ***Achievement 5:*** *Die Wahrnehmung von unangemessenen und hochwertigen Kommentaren potenzieller Diskussionsteilnehmenden ist umfänglich und sozial inklusiv in das Training unserer KI eingeflossen.*

Neben der Betrachtung der Ansichten der Moderator:innen sollte auch die Perspektive potentieller Diskussionsteilnehmenden in die Entwicklung von KOSMO einfließen. Zu diesem Zweck wurden im Rahmen des zweiten Praxistests ein Großteil der Trainingsdaten für die finale KOSMO KI mittels Crowdannotation durch formal unterschiedlich gebildete Clickworker erstellt. Hierdurch konnte zum einen eine für die wissenschaftliche Anschlussforschung bedeutsame Steigerung der Klassifikationsleistung der Moderations-KI sichergestellt werden. Zum anderen konnten so die Ansichten von potentiellen Nutzer:innen von KOSMO in Abgrenzung zur Perspektive der professionellen Moderator:innen über wünschenswerte und unangemessene Kommunikationshalte in Online-Diskussionen direkt in die Entwicklung der KI einfließen. Die Crowdannotation als in der Industrie übliches Verfahren zur Generierung von Trainingsdaten wurde für KOSMO auf Grundlage sozialwissenschaftlicher Befunde optimiert: Eine solide Studiengrundlage bestätigt, dass die Wahrnehmung von Inzivilität und deliberative Qualität entlang verschiedener sozialer Gruppen systematisch variiert. Entsprechend wurden Informationen über die Crowdannotierenden erhoben, um so relevante Kennwerte über die Datengrundlage ausweisen zu können sowie eine empirische Prüfung auf mögliche Verzerrungen innerhalb der Trainingsdaten der KI zu ermöglichen. Zusätzlich wurde erstmalig darauf geachtet, dass eine gleichmäßige Verteilung unterschiedlicher formaler Bildungsniveaus der Crowdannotierenden in den erhobenen Daten vorliegt. Während die Deliberationstheorie Bildungsniveaus als mögliche Verzerrungsquelle von

Deliberativitäts- und Inzivilitätswahrnehmung nahelegt, wurde eine dahingehende empirische Untersuchung bis jetzt noch nicht vorgenommen. Auf Grundlage der erhobenen Daten ist dies nun möglich. Das Forschungsvorhaben stieß bereits auf mehreren wissenschaftlichen Konferenzen auf reges Interesse, die Auswertung der Ergebnisse wird bis Ende des Jahres 2023 erfolgt sein.

4. Erfolgte und geplante Veröffentlichungen

Source Code/Software

Der Source Code der Plattform KOSMO wurde mit der ersten Iteration veröffentlicht. Das Repository ist unter folgendem Link auf git zu finden: https://github.com/liqd/a4-kosmo

Der Prototyp ist veröffentlich unter dem Link: https://kosmo.liqd.net/

Interviews und Berichterstattung

Mackisack, D. (2023). Exploring AI Moderation – a 'Kosmo' journey. *Democracy Technologies*. https://democracy-technologies.org/industry-news/exploring-ai-moderation-a-kosmo-journey/ (zuletzt abgerufen am 10.10.2023).

Mackisack, D. (2023b). Behind the scenes on open source democracy technology. *Democracy Technologies*. https://democracy-technologies.org/industry-news/behind-the-scenes-on-open-source-democracy-technology/ (zuletzt abgerufen am 10.10.2023).

EnergieAgentur.NRW. (2021, 4. Oktober). Bürgerbeteiligung mit Hilfe künstlicher Intelligenz. *Erneuerbare Energien - Der Podcast der EnergieAgentur.NRW.* https://www.podcast.de/episode/587060076/25-buergerbeteiligung-mit-hilfe-kuenstlicher-intelligenz (zuletzt abgerufen am 10.10.2023).

Wissenschaftliche Veröffentlichungen und Tagungsbeiträge

Fritz, J. (2021). *Sind wir dafür bereit, dass KI menschliche Moderation ersetzt? Ein interaktives Experiment zur Wahrnehmung von KI-Moderationsagenten in digitalen Debatten*. Unveröffentlichte Abschlussarbeit zur Erlangung des akademischen Grades Master of Arts, Philosophische Fakultät der Heinrich-Heine-Universität Düsseldorf.

Risch, J., Stoll, A., Wilms, L., & Wiegand, M. (2021). Overview of the GermEval 2021 shared task on the identification of toxic, engaging, and fact-claiming comments. In Risch, J., Stoll, A., Wilms, L., & Wiegand, M. (Eds.), *Proceedings of the GermEval 2021 Workshop on the Identification of Toxic, Engaging, and Fact-Claiming Comments: 17th Conference on Natural Language Processing KONVENS 2021* (pp. 1-12). Association for Computational Linguistics.

Stoll, A. & Heinbach, D. (2022). The More the Merrier? Training Strategies for AI in Algorithmic Content Moderation Systems. Vortrag auf der *72rd Annual Conference of the International Communication Association* (ICA), 26.-30. Mai 2023, Paris.

Stoll, A. & Heinbach, D., & Ziegele, M. (2022). Empirisch generierte Trainingsdaten für die KI-gestützte Moderation von Online-Diskussionen: Potenziale des Datentyps „Field-Labeled Data". Vortrag auf der *68. Jahrestagung der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPUK)*, 18.-20. Mai, Bremen.

Wilms, L., Gerl, K., Stoll, A. & Ziegele, M. (2022). What do you need from algorithmic transparency? Findings from Qualitative Interviews with moderators of online discussion fora in public administration and journalism. Vortrag auf der *9th European Communication Conference (ECREA 2022),* 19. - 22. Oktober 2022, Aarhus.

Wilms, L., Gerl, K., Stoll, A. & Ziegele, M. (2023). Technologieakzeptanz von und Transparenzanforderungen an automatisierte Hate-Speech Erkennung – Erkenntnisse aus qualitativen Interviews mit Moderator:innen von

Online-Diskussionen öffentlich-rechtlicher Medien und Verwaltung. Vortrag auf der *68. Jahrestagung der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPUK)*, 18.-20. Mai, Bremen.

Wilms, L., Stoll, A. & Ziegele, M. (2023). Uncovering educational bias in crowd-annotated data for automated classification of constructive and uncivil comments. Vortrag auf der *73rd Annual Conference of the International Communication Association (ICA)*, 25.-29. Mai 2023, Toronto.

Wilms, L., Stoll, A., Ziegele, M. & Gerl, K. (2023). Bildungsbezogene Biases in crowd-annotierten Daten zur automatischen Klassifikation von konstruktiven und inzivilen Kommentaren. Vortrag auf der *Gemeinsamen Jahrestagung der Fachgruppe "Kommunikation und Politik" der DGPUK, des Arbeitskreises "Politik und Kommunikation der DVPW und der Fachgruppe "Politische Kommunikation" der SGKM,* 28.-30. Juni, Düsseldorf

Wilms, L., Gerl, K., Stoll, A. & Ziegele, M. (2023). Technology Acceptance and Transparency Demands for Automated Detection of Toxic Language – Interviews with Moderators of Public Online Discussion Fora. *Journal of Human-Computer Interaction.* (under review)

In Planung: Wissenschaftliche Publikation der KOSMO KI (AP2)

In Planung: Wissenschaftliche Publikation des zweiten Praxistests (AP6, AP2)

Vorträge

Seim, J., Wilms, L., Fröhlingsdorf, S., Heinbach, D. & Siemer, M.K. (2022). Panel Jenseits von Blockieren und Löschen. *Online-Kongress zur digitalen Demokratie D3: #Deutschland #Digital #Demokratisch*. Berlin Institut für Partizipation, 22.-23. November 2023, digital.

Siemer, M.-K. & Wegener, F. (2022). AI for democracy! Wie KI demokratische Diskurse unterstützen kann (Tool Inside). *Online-Kongress zur digitalen Demokratie D3: #Deutschland #Digital #Demokratisch*. Berlin Institut für Partizipation, 22.-23. November 2023, digital.

Stoll, A. & Siemer, M.K. (2023). High Standards and Low Budget – Künstliche Intelligenz für demokratische Öffentlichkeiten. *re:publica*, 05.-07. Juni 2023, Berlin.

Siemer, M.-K. (Okt 2022). Vorstellung von KOSMO im Ausschuss Bürger*innenbeteiligung Pankow.

In Planung: Wehking, R R. (2023). Wie kann Künstliche Intelligenz, in der Anwendung von KOSMO, bei der Moderation deutschsprachiger Online-Diskussionen unterstützen? *Symposium „Digitale Öffentlichkeitsarbeit" der Bundeswehr.*