



Institut für
Medienforschung und
Medienpädagogik

Technology
Arts Sciences
TH Köln



CB MANAGEMENT GmbH

Anke Hildebrandt | Paulina Roloff | Marina Schubert | Friederike Siller

Co-Constructing KI – CoCoKI

Entwicklungslabor für einen geschützten
Kommunikationsraum für Kinder. (*Working Paper*).

Impressum:

Autorinnen: Friederike Siller, Anke Hildebrandt, Paulina Roloff, Marina Schubert
Technische Hochschule Köln
Institut für Medienforschung und Medienpädagogik
Fakultät für Angewandte Sozialwissenschaften
Gustav-Heinemann-Ufer 54
D-50968 Köln

Kontakt:

Prof. Dr. Friederike Siller
friederike.siller@th-koeln.de

Köln, Oktober 2024

Kerndaten zum Projekt	
Projekttitle	Co-Constructing KI – CoCoKI, Entwicklungslabor für einen geschützten Kommunikationsraum für Kinder
Projektziel	Bausteine für eine Grundkonzeption kindgerechter sozialer Medien mit Schwerpunkt auf dem Einsatz von Algorithmen
Zeitraumen	September bis November 2023
Workshop 1	Konzeptions-Workshop Oktober 2023
Workshop 2	Test- und Feedback-Workshop zum Prototyp November 2023
Feldzugang	Konfirmandengruppe der Kartäuserkirche Köln
Verbundpartner und Beteiligte	<p>Seitenstark e. V. (Verbundleitung) Helga Kleinen, Anke Hildebrandt*, Paulina Roloff*</p> <p>Institut für Medienforschung und Medienpädagogik der Technischen Hochschule Köln (IMM) Prof. Dr. Friederike Siller*, Franziska Schäfer, Marina Schubert*, Hannah Rodenbüsch</p> <p>Gemeinschaftswerk der Evangelischen Publizistik gGmbH Rahel Kleinwächter, Dieter Schmitt, Moritz Vogel CB Management GmbH Christoph Becker, Dr. Klaus Holthausen</p> <p><i>* Autorinnen des Kurzberichts</i></p>
Gefördert durch	<p>Bundeszentrale für Kinder- und Jugendmedienschutz</p> <p>Im Rahmen des Förderprogramms: „Weiterentwicklung des Kinder- und Jugendmedienschutzes“ (2023)</p>

Hintergrund zum Projekt

Das Projekt „CoCoKI – Co-Constructing KI“ widmete sich der Frage, wie ein altersgerechter und geschützter Kommunikationsraum für Kinder im Alter von neun bis 13 Jahren ausgestaltet sein sollte, um Anforderungen und Bedürfnissen des Kinder- und Jugendmedienschutzes zu erfüllen. Gemeinsam mit Kindern, Medienpädagoginnen und Entwicklern wurden dazu Ansätze erarbeitet und getestet.

Für das Projektvorhaben ergaben sich leitende Fragestellungen:

- Welche Anforderungen und Bedarfe haben Kinder an Kommunikationsräume in Bezug auf Inhalte und Funktionen?
- Welche „kindgerechten Algorithmen“ für soziale Medien lassen sich identifizieren und wie lassen sich diese gestalten?
- Wie gelingt eine ko-konstruktive Design- und Produktentwicklung eines Kommunikationsraumes in einem Workshopformat?

Für die Bearbeitung dieser Fragen wurde ein innovativer Ansatz gewählt, in dem Kinder und Entwickler gemeinsam in einem ko-konstruktiven Prozess (in „Entwicklungslaboren“) an der Konzeption und Prototyp-Entwicklung eines sozialen Netzwerks für Kinder gearbeitet haben. Der Feldzugang für die Beteiligung der Kinder konnte über eine Kölner Gemeinde der evangelischen Kirche realisiert werden. Die beteiligten Kinder waren durchschnittlich 13 Jahre alt und damit an der Obergrenze der definierten Zielgruppe der neun- bis 13-Jährigen. Alle teilnehmenden Kinder verfügten bereits über vielfältige Erfahrungen mit sozialen Netzwerken und Messengern wie Snapchat und WhatsApp.

In vier Projektphasen wurden Anliegen, Interessen und Wünsche von Kindern an digitale Kommunikationsräume identifiziert, Ideen für kindgerechte Algorithmen und Funktionen erarbeitet sowie in Teilen prototypisch umgesetzt und gemeinsam mit den Kindern getestet. Das Format einer solchen ko-konstruktiven Ideen-, Konzept- und Produktentwicklung birgt viel Potenzial, um innovative, neue Ansätze und Ideen hervorzubringen. Durch den Einsatz vielfältiger Methoden und dem Bewusstsein für eine kindgerechte Ansprache kommen Kinder und Entwickler direkt und ohne „kommunikative Reibungsverluste“ miteinander in Kontakt.

Aus den Beiträgen der Kinder in Workshop 1 ließen sich sechs wesentliche Aspekte ableiten, die für ein kindgerechtes soziales Netzwerk bzw. eine Kommunikationsplattform von zentraler Bedeutung sind. Diese waren:

- 1) Effektive Filterfunktionen und anbieterseitige Reaktionen auf Hassrede
- 2) Hoher Grad an Entscheidungsmacht und Selbstbestimmung bei den Nutzenden
- 3) Niedrigschwellige Funktionen zum Teilen von Inhalten, Veröffentlichen von Fotos und Videos, Chatfunktion
- 4) Dashboard („Alles in einem zu haben“)
- 5) Autonomes Zeitmanagement)
- 6) Kindgerechte Algorithmen und Transparenz

Auf Basis der Ergebnisse aus Workshop 1 initiierte das Team eine erste technische Implementierung des Prototyps in Phase 2. Dabei lag der Fokus auf dem Aufsetzen des Basissystems, das zunächst die Einrichtung der Server-Instanz mit der Rohfassung des Netzwerks, einschließlich relevanter Parameter, Basisalgorithmen und eines Hate-Speech-Filters umfasste. Parallel dazu wurden circa 40 Demo-Nutzerkonten mit Comic-Avataren erstellt, um eine Grundlage für die Implementierung neuer Funktionen und allgemeiner Features zu schaffen.

Um die Rohfassung gestalterisch für die Kindergruppe des Projekts (Konfirmandengruppe) aufzubereiten, wurden exemplarisch Icons aus dem kirchlichen Umfeld implementiert (z.B. Lutherrose als Gefallen-Button).

Das gemeinsame Testen des Prototyps wurde in Workshop 2 vorgenommen. Dabei konnten die vielfältigen Rückmeldungen der teilnehmenden Kinder sowie die umfassende Berücksichtigung einer jüngeren Zielgruppe maßgeblich zur Verfeinerung und Ausrichtung der Prototyp-Konzeption beitragen.

Die Workshops bildeten demnach die Grundlage für die (Weiter-)Entwicklung eines Prototyps. Dabei standen die Wünsche und Anliegen der Kinder sowie Kinder- und Jugendschutzmaßnahmen im Fokus. Es wurden im Prozess gezielt die Funktionen identifiziert und dazu weiter konzeptionell wie technisch gearbeitet, die im Rahmen des Projektzeitraums technisch umsetzbar und insbesondere für die Kinder von Bedeutung waren: eine selbstbestimmte, autonome Nutzung (z.B. über Zeiteinstellungen auf der Plattform), ein angemessener Umgang mit Hate-Speech inkl. anbieterseitigen Reaktionen, die Möglichkeit, die eigenen Emotionen über die

Plattform angemessen zu artikulieren sowie der Wunsch nach Möglichkeiten der individuellen Profilgestaltung. Welche Elemente auf Basis der Workshop-Ergebnisse im Ansatz umgesetzt wurden und welche perspektivisch gedacht werden können, zeigt die Tabelle im Anhang. Vorrangige Zielsetzung des Projekts bestand darin, Kinder aktiv an der Gestaltung eines kindgerechten Kommunikationsraums zu beteiligen und dabei gemeinsam neue Ansätze, Ideen und Elemente für die Konzeption eines Netzwerks zu entwickeln, ohne jedoch eine abschließende und voll funktionsfähige Version bereitzustellen.

Die ko-konstruktive Gestaltung ermöglichte dabei einen kontinuierlichen Lernprozess auf allen Seiten, bei dem Ideen nicht nur generiert (Workshop 1), sondern auch in der Praxis erprobt wurden (Workshop 2).

Es lässt sich festhalten, dass eine ansprechende Gestaltung und eine leicht verständliche Nutzung den Kindern besonders wichtig sind. Altersangemessene Sicherheitsfunktionen, vor allem für jüngere Kinder, und Elemente zur Förderung von Fairness in digitalen Räumen, sind entscheidende Schlüsselpunkte für die Ausgestaltung eines sicheren Kommunikationsraums für Kinder. Daneben waren ein vertieftes Verständnis der verwendeten Algorithmen und eine diesbezügliche Transparenz der Anbieter wichtige Anliegen der Kinder.

Kindgerechte Mechanismen und Algorithmen

Von den Kindern und dem Entwicklungsteam wurden gemeinsam konzeptionelle Überlegungen und anwendungsnahe Vorschläge für ein faires soziales Netzwerk für Kinder ausgearbeitet. Zu den zentralen Umsetzungsideen auf technischer Ebene, die in dem Projekt erprobt wurden, gehören die folgenden kindgerechten Mechanismen und Algorithmen:

1. **Aufmerksamkeits-Algorithmus:** *Wer andere Inhalte liked und Nutzenden damit Aufmerksamkeit schenkt, wird vom Algorithmus höher eingestuft bzw. belohnt und damit im Empfehlungssystem stärker berücksichtigt. (Umdeutung des klassischen Belohnungs-Prinzips der Empfehlungsalgorithmen, die nach einem Pyramidensystem funktionieren).*
2. **Fairness-Algorithmus:** *Wer auf unfaire und unangemessene Sprache verzichtet und positive Sprache verwendet, wird höher eingestuft. Wer hingegen unangemessene Sprache, z.B. Schimpfwörter verwendet, wird heruntergestuft. Grundlage des Algorithmus ist ein Wortfilter auf Basis einer manuell angelegten **Blacklist**.*

3. **Empfehlungs-Algorithmus:** *Das System erkennt, wer ähnliche Interessen und Hobbys hat und kann gemeinsame Themenfelder ausspielen. (Das typische Follower*innen-System, also beispielsweise Kontaktlisten und Abonnements, wird ignoriert. Stattdessen werden Gleichgesinnte identifiziert, indem übereinstimmende Likes aller Kategorien gewertet werden).*
4. **Entity-Recognition:** *Das System erkennt nicht nur, wer ähnliche Interessen hat aufgrund von gleichen „Gefällt mir“-Angaben, sondern nutzt ebenfalls eine Inhaltserkennung („Entity-Recognition“), um Beiträge zu einem zuvor festgelegten Themenfeld zu identifizieren und gesammelt auf dem Dashboard anzuzeigen. (Im Prototyp ist die Wortliste auf den Themenbereich „Kirche“ ausgelegt, um dem Kontext der teilnehmenden Kinder (Konfirmandengruppe) zu entsprechen).*
5. **Mitgefühl-Algorithmus:** *Die emotionale Bewertung von Beiträgen ermöglicht ein unterstützendes Umfeld. Verwendet ein Kind überwiegend traurig-konnotierte Wörter in seinen Beiträgen, summieren sich diese zu einem entsprechenden Score. Kinder mit betrübter Stimmungslage werden identifiziert. Andere Nutzer*innen können daraufhin ein Zeichen der Unterstützung abgeben.*
6. **Fairscore (kombinierter Algorithmus):** *Die Algorithmen-Kombination funktioniert als Anreiz und Würdigung für positives Verhalten. Sie berücksichtigt und gewichtet fünf verschiedene Faktoren: Aufmerksamkeits-Score, Fairness-Score, MitgefühlScore sowie der Verwarnungs-Score (steigt bei Eingabe unangebrachter Sprache) und Meldungen via Melde-Button durch andere Nutzende.*

Die im Projekt durchgeführte Patentrecherche im Bereich der Empfehlungsalgorithmen hat ergeben, dass eine Weiterentwicklung der Algorithmenkonzepte in Deutschland gegebenenfalls patentierbar wäre. Eine weitere Befassung mit dem Konzept sollte Fachdisziplinen wie Psychologie, Kindheitspädagogik, Ethik und Kommunikationswissenschaften miteinbeziehen, um den Einsatz im Kinderumfeld vorzubereiten.

Im Kontext konzeptioneller Überlegungen innerhalb des Projekts wurden weitere Aspekte berücksichtigt:

Nutzungssetting (formal, non-formal, informell)

Immer wieder wurde deutlich, dass der Nutzungskontext einer sozialen Kommunikationsplattform für Kinder entscheidend für die konkrete Ausgestaltung von Sicherheitsmechanismen und Funktionen ist (Nutzung z.B. im schulischen,

außerschulischen oder informellen Bereich). So stellt sich die Frage, ob der Kommunikationsraum öffentlich zugänglich sein soll und von allen Kindern im Internet genutzt werden kann, oder ob er dem Austausch einer geschlossenen Kindergruppe dient, beispielsweise im Konfirmandenunterricht.

Einsatz von künstlicher Intelligenz

Der gezielte Einsatz von künstlicher Intelligenz (KI) für einen Kommunikationsraum für Kinder kann neue Möglichkeiten eröffnen. In dem Prototyp kommt KI zur Anwendung, um automatisierte und intelligente Funktionen zu ermöglichen. KI bezieht sich dabei auf Algorithmen, die in der Lage sind, Informationen zu analysieren, Muster zu erkennen und Entscheidungen zu treffen. Weitergehende Funktionen und Möglichkeiten von KI umzusetzen, kann Aufgabe eines Folgeprojektes sein. Überlegungen, welche dies sein könnten, wurden im Projekt eruiert: Überlegt wurde etwa, dass es durch KI in dem entwickelten Kommunikationsraum möglich sein soll, den Kontext von Eingaben der Nutzenden (sowohl Text als auch Bild) zu erkennen, um gegebenenfalls entsprechende Filter zu aktivieren. Wäre ein intelligenter **Wortfilter** zukünftig in der Lage, zuverlässig und weitestgehend objektiv Hate-Speech eigenständig und dynamisch zu identifizieren, könnte menschliche Moderation entlastet werden.

Ko-Konstruktion

Über das Projekt wurde Raum für offenen Dialog zwischen Zielgruppe und Entwicklungsteam geschaffen. Dabei konnten gemeinsam vielversprechende Konzeptideen erarbeitet und ansatzweise umgesetzt sowie getestet werden.

Der Austausch zwischen Kindern und Entwicklern stellte sich als durchaus anspruchsvoll und voraussetzungsvoll dar. Medienpädagogische Moderation und „Übersetzung“ waren erforderlich, um die unterschiedlichen Sichtweisen und Perspektiven zusammenzubringen, eine gemeinsame Sprache zu finden und für gegenseitiges Verständnis und Vertrauen zu sorgen.

Die Einbindung der Kinder in den Entwicklungsprozess erwies sich insgesamt als sowohl konstruktiv als auch erkenntnis- und aufschlussreich. Sie waren reflektiert, interessiert und motiviert, ihre Bedürfnisse zu äußern und eigene Ideen einzubringen. Dabei ging es ihnen nicht nur um das Einbringen neuer Features und Designs, sondern auch um die grundlegende Konzeption eines geschützten und selbstbestimmten Kommunikationsraums, der ihren Bedürfnissen entspricht.

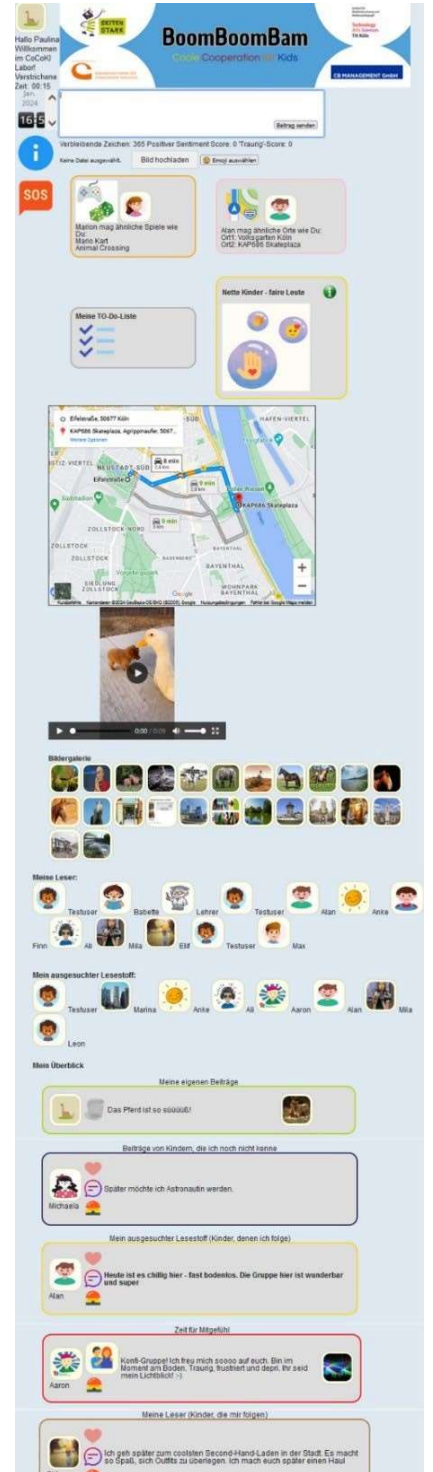
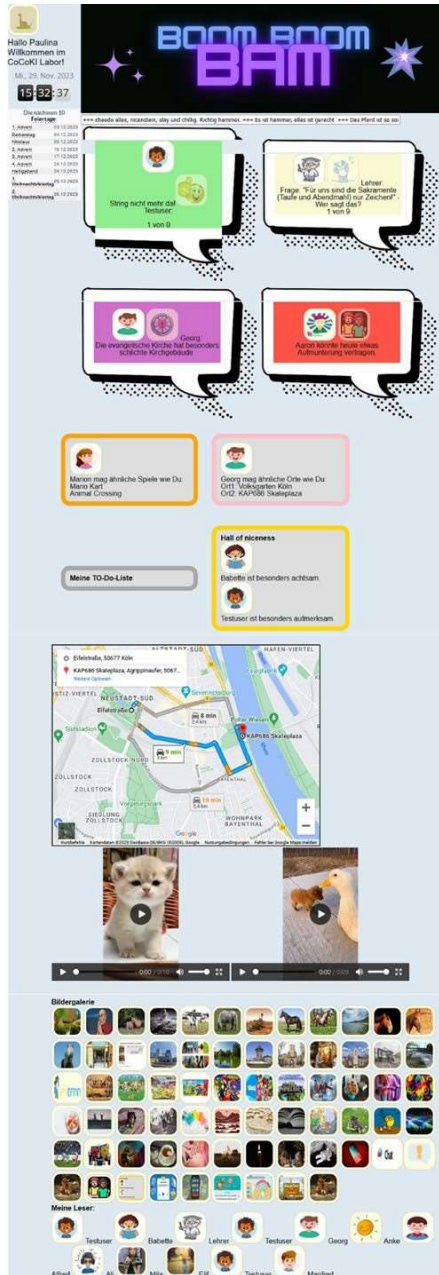
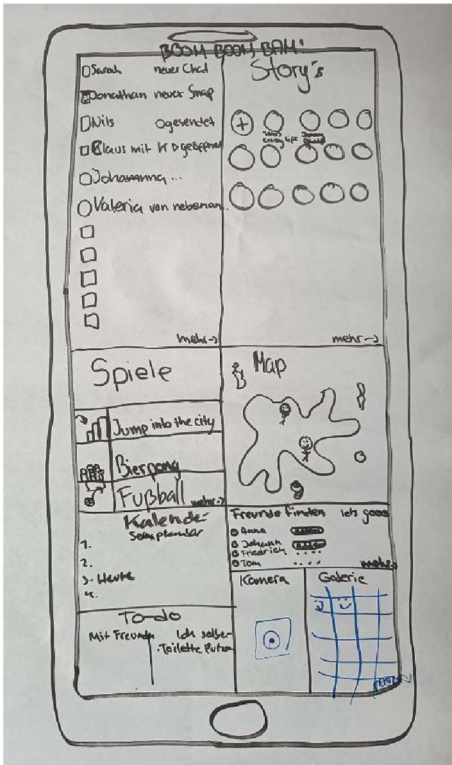
Anbindung an die Kinderseitenlandschaft

1. Bilddatenbanken: Mögliche Anbindung an bereits geprüfte Bild-Datenbanken. Kooperationen wären je nach Ausgestaltung z.B. denkbar mit Find-das-Bild.de, der BilderSuche von FragFINN.de oder der Kinder-Foto-Community Knipsclub.de.
2. Newsfeed: Ein im Prototyp exemplarisch eingebundener Newsfeed der SeitenstarkKinderseiten entspricht dem Wunsch nach aktuellen Nachrichten und Vernetzung.
3. Kontrolle über eingefügte Links/URLs sowie Videos: Mit einer Filterliste und den geprüften Angeboten der Kinderseiten wären sichere URLs sowie kindgerechte Videoangebote automatisiert einzubinden.

Abschließend angemerkt sei, dass die erarbeiteten Erkenntnisse, von der Gestaltung und Benutzerfreundlichkeit bis hin zu Sicherheitsaspekten und Vernetzungsmöglichkeiten, eine gute Grundlage für eine weitere Bearbeitung des Themas bilden. Die Projektergebnisse beinhalten **transferierbare Impulse** für die Entwicklung von Kinder-Onlineformaten unter Beteiligung junger Menschen, und für innovative Konzeptionen von Kommunikationsräumen für junge Nutzende.

Eine weitere Befassung mit und ein weiteres Voranbringen der im Projekt erarbeiteten Ergebnisse würde einen vielversprechenden und wichtigen Beitrag für den Jugendmedienschutz und die Kindermedienlandschaft darstellen. Hierfür sind Ressourcen und eine längere Projektlaufzeit von wesentlicher Bedeutung (z.B. für die Kalibrierung, Optimierung der Algorithmen, den Design-Prozess mit Kindern, die Evaluation, den Einbezug disziplinärer Perspektiven).

Beispiele und Screenshots aus dem Arbeitsprozess



Elemente und Bausteine für ein kindgerechtes soziales Netzwerk

Im Ansatz umgesetzte Elemente/Bausteine		Perspektivische Elemente/Bausteine
Filterfunktion für Hate-Speech	Selbstbestimmte Zeiterinnerung	Filterfunktion für Bildinhalte
Meldefunktion: Meldebutton	Kindgerechte Erklärungen zur Funktionalität von Algorithmen	Kontexterkennung
Dialogboxen zur Sicherung von Transparenz & Verständlichkeit	Hilfefunktion: Hilfebutton	Spamfilter
Feedback-Boxen	Widgets & interaktive Features	Individuelle Anpassung von Privatsphäre-Einstellungen
Selbstbestimmte Zusammenstellung von Funktionalitäten	Newsfeed	Individualisierung des Accounts
Veröffentlichung von Bildern und Videos	Aufmerksamkeits-Algorithmus	Direktnachrichten
Eingabefeld mit Emoji-Auswahl	Fairness-Algorithmus	KI-gestützte Rechtschreib-Korrektur
allgemeingültige Stimmungsindikatoren	Empfehlungs-Algorithmus	Kommentarfunktion
Multifunkt. Nutzeroberfläche (Dashboard)	Entity-Recognition	Allround-Suchfunktion
Persönliche Timeline	Mitgefühl-Algorithmus	Definition von Rollen
Gefällt-mir-Markierung	Fairscore	Niederschwelliger Log-In
		Mobile Anwendung
		Anbindung an geprüfte Bild-Datenbanken
		Kontrolle über eingefügte Links/URLs
		Einbindung von kindgerechten Videos