

Strategien und Werkzeuge zur medienübergreifenden Meldung, Erkennung und Behandlung von Cybermobbing und Hassbotschaften

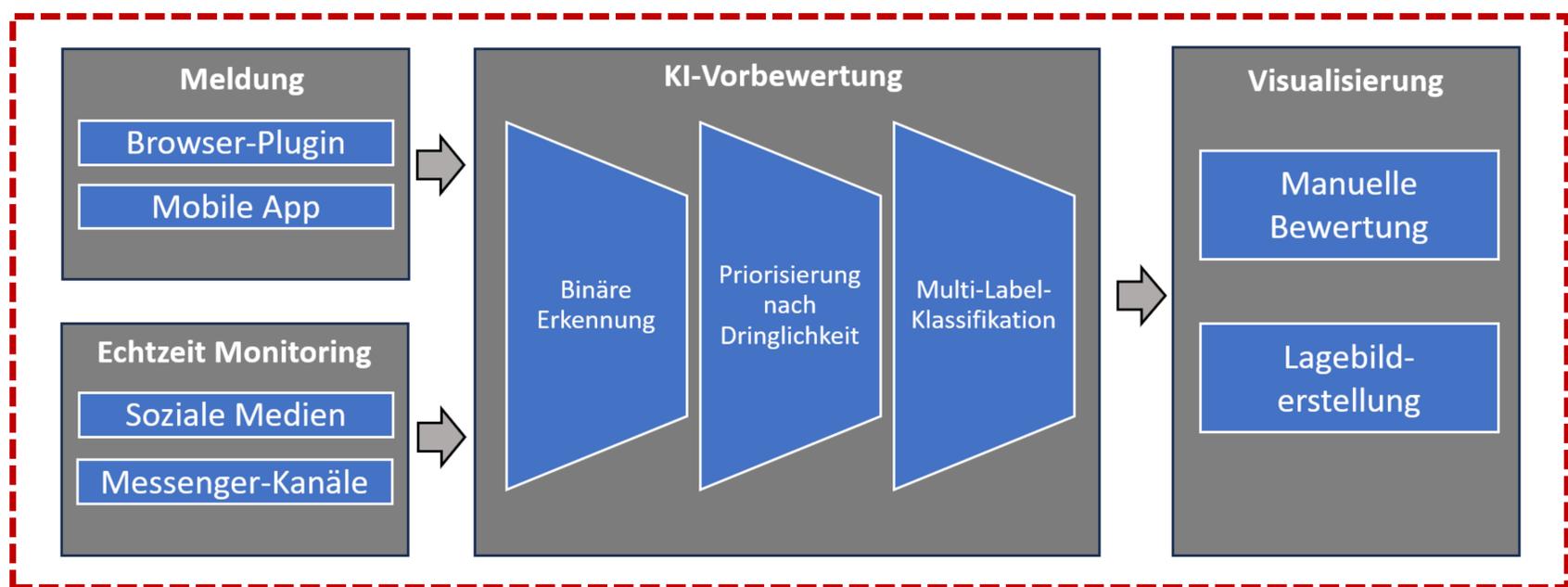
Julian Bäumlér ♦ Marc-André Kaufhold ♦ Christian Reuter ♦
Stefan Stieglitz ♦ Milad Mirbabaie ♦ Christoph Fuchß

Ziel des Projekts

- Unterstützung von **Ermittlungs- und Strafverfolgungsbehörden** sowie **Meldestellen** durch:
 - Neuartige Strategien und Werkzeuge zur Meldung, Erkennung und Behandlung von Cybermobbing und Hassbotschaften
- Ein neuartiger Demonstrator unterstützt die:
 - **Integration gemeldeter Vorfälle** per App, Browser-Plugin, E-Mail oder Webformular
 - Echtzeitbasierte **Sammlung** von Vorfällen in sozialen Datenquellen (z.B. öffentliche Messenger-Kanäle und soziale Medien)
 - Konfigurierbare und visuelle **Analyse** für die Früherkennung und Priorisierung auf Basis eines Dashboards

Angestrebte Innovationen

- Taxonomie** zur Klassifikation und Behandlung von missbräuchlichen Inhalten in sozialen Medien
- Methoden** zur anpassbaren, fairen und nachvollziehbaren Klassifikation von Cybermobbing und Hassbotschaften
- Demonstrator** zur medienübergreifenden Meldung, Erfassung und Analyse von Cybermobbing und Hassbotschaften
- Organisationale Strategien** zur systematischen intra- und interorganisationalen Analyse von und Kommunikation zu missbräuchlichen Inhalten



Forschungspotentiale aus der Praxis

- (1) Ereignis- und akteurszentrierte Detektion von Hassbotschaften
- (2) Priorisierung von Hassbotschaften nach Dringlichkeit
- (3) Nachvollziehbare Vorbewertung von Hassbotschaften, vor allem nach strafrechtlicher Relevanz
- (4) Unterstützung der kollaborativen und interorganisationalen Fallbearbeitung
- (5) Barrierefreie und sichere Meldung und Dokumentation von Hassbotschaften
- (6) Teilautomatisierte Lagebilderstellung und Visualisierung

Aktuelle Schwerpunkte

- Anforderungen und Bedenken Jugendlicher bei der Meldung
- Browser-Plugin zur Erkennung, Dokumentation und Meldung
- Hassbotschaften im Kontext von Desinformation
- Nachvollziehbare Multi-Label-Klassifikation von Hassbotschaften
- Gesellschaftliche Akzeptanz behördlicher KI-Nutzung

Projektname: Entwicklung von Strategien und Werkzeugen zur medienübergreifenden Meldung, Erkennung und Behandlung von Cybermobbing und Hassbotschaften in Ermittlungs- und Strafverfolgungsbehörden (CYLENCE)

Schlüsselwörter: Cybermobbing, Hassbotschaften, Meldestelle, Lagebilderfassung, Ermittlungs- und Strafverfolgungsbehörden, Social Media Analytics, Künstliche Intelligenz, Visual Analytics

Förderung: Programm „Zivile Sicherheit - Bedrohungen aus dem digitalen Raum“ des Bundesministeriums für Bildung und Forschung (BMBF)

Förderkennzeichen: 13N16636 bis 13N16639
Laufzeit: 08.2023 – 07.2026
Projekträger: VDI Technologiezentrum GmbH

Ansprechpartner

Julian Bäumlér
Wissenschaftlicher Mitarbeiter CYLENCE
E-Mail: baeumler@peasec.tu-darmstadt.de

Dr. Marc-André Kaufhold
Projektmanager CYLENCE
E-Mail: kaufhold@peasec.tu-darmstadt.de

Prof. Dr. Dr. Christian Reuter
Verbundkoordinator CYLENCE
E-Mail: reuter@peasec.tu-darmstadt.de

Technische Universität Darmstadt
Wissenschaft und Technik für Frieden und Sicherheit (PEASEC)
Pankratiusstraße 2, 64289 Darmstadt
Website: <https://peasec.de/>

