CY-WARN
Cyber Situational Awareness & Communication

Abschlussbericht des BMBF-Projekts „CYWARN“

Zuwendungsempfänger:	**Technische Universität Darmstadt (TUDa) Wissenschaft und Technik für Frieden und Sicherheit Prof. Dr. Dr. Christian Reuter Pankratiusstraße 2 64289 Darmstadt**
Förderkennzeichen:	**13N15407**
Verbundtitel:	**Strategie- und Technologie-Entwicklung zur medienübergreifenden Erstellung eines Cyber-Lagebilds und akteursspezifischen Kommunikation von Cyber-Warnmeldungen (CYWARN)**
Teilvorhabentitel:	**Methoden, Strategien und Technologien zur medienübergreifenden und teilautomatisierten Cyber-Lagebilderstellung**
Laufzeit des Vorhabens:	**01.10.2020 – 31.03.2024**
Berichtsverfasser:	**Dr. Marc-André Kaufhold**
Erstellungsdatum:	**16.09.2024**

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SIFO.de

TECHNISCHE UNIVERSITÄT DARMSTADT

1 Verwendung der Zuwendung

Das Gesamtziel von CYWARN ist es, für Computer Emergency Response Teams (CERTs) Strategien und Technologien zur Analyse und Kommunikation des Cyber-Lagebilds zu entwickeln. Dies trägt dazu bei, die Früherkennung von Cybergefahren (Vorbereitung) und Umsetzung von Gegenmaßnahmen (Reaktion) zum Schutz von Medien- und Informationsinfrastrukturen durch CERTs zu verbessern. Der Prozess wird durch einen neuartigen Demonstrator unterstützt und umfasst (I) die automatisierte Erhebung und Integration öffentlicher und geschlossener Datenquellen, (II) die intra- und interorganisationale Datenauswertung mit Glaubwürdigkeitsanalyse der Inhalte und Quellen unter Zuhilfenahme von Visual Analytics und (III) die kanalübergreifende Kommunikation von Cyberbedrohungen an verschiedene Akteure (BürgerInnen, Behörden und KRITIS-Betreiber), um diese zum bestmöglichen Eigenschutz zu befähigen. Der Arbeitsplan umfasst theorieorientierte Literaturstudien, qualitative und quantitative empirische Studien (AP1), partizipative Demonstrator- und Strategie-Entwicklung (AP2, AP3), Evaluationen der Akzeptanz und Gebrauchstauglichkeit (AP4) und eine Abschätzung der ethischen, rechtlichen und sozialen Implikationen (AP5).

	Analyse	**Kommunikation**
Strategie	Entwicklung von Analysestrategien (PEASEC/RLP)	Entwicklung von Kommunikationsstrategien (Hessen3C)
Technologie	Entwicklung von Analysetechnologien (PEASEC)	Entwicklung von Kommunikationstechnologien (PROCO)
Integration	Entwicklung des integrierten Gesamtdemonstrators (VAG)	

Tabelle 1: Schwerpunktmatrix in CYWARN, in grün: PEASEC-Schwerpunkt

Das Teilvorhaben „Methoden, Strategien und Technologien zur medienübergreifenden und teilautomatisierten Cyber-Lagebilderstellung“ verfolgte das Ziel, unter Einbezug automatisierter Methoden CERTs in der Erhebung und Auswertung von Cyberbedrohungen und Sicherheits-lücken zur verbesserten Cyber-Lagebilderstellung zu unterstützen. Da technologische Entwicklungen im Rahmen der Digitalisierung, insbesondere dem Internet der Dinge und Smart Cities, für eine immer stärkere Vernetzung in privaten und professionellen Kontexten sorgen, erhöht sich ebenso die Gefahr durch Cyberbedrohungen und Sicherheitslücken in Medien- und Informationsinfra-strukturen. Zur Früherkennung von Cyberbedrohungen und Sicherheitslücken sind CERTs mit einer steigenden Anzahl öffentlicher Informationen und Quellen (z.B. Blogs, Feeds, Foto- und Videoportale, Sensoren, soziale Netzwerke und Websites) konfrontiert, die aufgrund unterschiedlicher Informationsqualität und Quellenvertraulichkeit zur Integration in ein Cyber-Lagebild gefiltert, analysiert, verifiziert und anschließend priorisiert werden müssen. Die zentralen Ziele des Teilvorhabens sind dementsprechend:

1) Die empirische Erfassung von aktuellen Arbeitspraktiken und Anforderungen von CERTs im Hinblick auf die geplante Strategie- und Technologie-Entwicklung sowie der Erwartungshaltung und Wahrnehmung von BürgerInnen (AP1).
2) Identifikation und Konzeption der automatisierten Erfassung öffentlicher Datenquellen (Open Source Intelligence) zur Integration in ein Cyber-Lagebild (Cyber Situational Awareness) (AP2).
3) Die Entwicklung von Methoden und Technologien auf Basis von Indikatoren und künstlicher Intelligenz sowie Strategien zur Filterung, Bewertung und Priorisierung von Cyber-Lageinformationen (AP3).
4) Die Integration und Verwertung der entwickelten und evaluierten Strategien und Technologien zur Etablierung des Cyber-Lagebilds in CERTs, die Veröffentlichung wissenschaftlicher Ergebnisse und der Transfer für andere Anwenderzielgruppen (AP4).
5) Die Durchführung einer ethischen und sozialen Technikfolgenabschätzung sowie die Analyse rechtlicher und organisationaler Bedingungen und Implikationen (AP5).

1.1 Stand der Forschung und Analyse des Nutzungskontextes (AP1)

Das Ziel von AP 1 war die Schärfung der theoretischen Fundierung und die Erhebung von Nutzeranforderungen für die Strategie-Entwicklung sowie an das Basis-, Analyse- und Kommunikationsmodul des zu entwickelnden Demonstrators basierend auf einer systematischen Literaturanalyse sowie qualitativen und quantitativen Untersuchungen.

1.1.1 Literaturanalyse: Herausforderungen und Potenziale der Forschung und Praxis (AP1.1)

Der Begriff Cyber-Situationsbewusstsein (Cyber Situtional Awareness) beschreibt ein Verständnis, das es Einzelpersonen ermöglicht, relevante Elemente in der Cyber-Umgebung innerhalb eines bestimmten Zeitrahmens und räumlichen Kontexts wahrzunehmen, ihre Bedeutung zu interpretieren und ihren zukünftigen Status vorauszusehen. In der Literatur wird dabei häufig zwischen Netzwerk- (Network), Bedrohungs- (Threat) und Missionsbewusstsein (Mission Awareness) unterschieden.

Netzwerkbewusstsein: In den meisten Fällen besteht die Hauptaufgabe der CERTs darin, ihre jeweilige Behörde, ihr Unternehmen oder ihre Organisation vor Cyberangriffen zu schützen, was ein Bewusstsein für die von ihnen genutzte Infrastruktur und ihre Netzwerke erfordert. Die Schaffung eines Netzwerkbewusstseins erfordert daher ein Verständnis der Struktur von Organisationsnetzwerken, die Verwaltung aller Anlagen und Konfigurationen, ein robustes Patch- und Upgrade-Management sowie eine routinemäßige Überprüfung von Schwachstellen, um eine Ausnutzung zu verhindern. Dementsprechend haben sich in der Unternehmenspraxis mehrere Kategorien von Werkzeugen etabliert, wie z.B. das Sicherheitsinformations- und Ereignis-Management (SIEM) für die Erkennung, Aggregation und Echtzeitanalyse sicherheitsrelevanter Informationen sowie Intrusion Detection and Prevention Systems (IDPS) zur Verhinderung von Schäden für Unternehmen oder die Gesellschaft durch Cyberangriffe.

Bedrohungsbewusstsein: Dennoch müssen die CERTs einen Überblick über die allgemeine Bedrohungslandschaft haben, um die Sicherheitsmaßnahmen der Organisationen anzupassen und Dienstleistungen für externe Kunden zu erbringen. Dies erfordert die Identifizierung verdächtiger Verhaltensweisen, aktuelle Informationen über externe Bedrohungen und die Teilnahme an Informationsaustauschgemeinschaften, um über neue und aufkommende Bedrohungen auf dem Laufenden zu bleiben. Im Hinblick auf Schwachstellen ist die National Vulnerability Database (NVD) ein Beispiel, in der Common Vulnerabilities and Exposures (CVEs) für die Klassifizierung von Schwachstellen verwendet werden und das Common Vulnerability Scoring System (CVSS) zur Bewertung des Schweregrads von CVEs anhand von Basis-, Zeit- und Umgebungsmetriken eingesetzt wird. Darüber hinaus geben zahlreiche Hersteller über ihre eigenen Kommunikationskanäle Informationen zu Schwachstellen und Lösungen heraus. Diese Sicherheitshinweise informieren über neue Sicherheitslücken und aktuelle Sicherheitsupdates für ihre Produkte. Darüber hinaus wurden Bedrohungsinformationsplattformen eingerichtet, um die Sammlung und Analyse von Daten über Cyber-Bedrohungen, z.B. Kompromittierungshinweise (Indicators of Compromise), zu erleichtern.

Missionsbewusstsein: Schließlich umfasst das Missionsbewusstsein eine angemessene Reaktion auf Sicherheitsereignisse, die Einstufung oder Priorisierung von Sicherheitsvorfällen im Hinblick auf ihre Auswirkungen auf den Auftrag oder das Geschäft der Organisation und die Vorwegnahme von Bedrohungen und Risiken durch die Durchführung von Risiko- und Bereitschaftsbewertungen. Die Arbeit der CERTs ist anspruchsvoll, da sie eine effektive Ausbildung, Koordination und den Austausch von Informationen sowohl innerhalb des Teams als auch mit externen Stellen erfordert. In Deutschland wird die Zusammenarbeit zwischen Bundes-, Landes- und privaten CERTs als entscheidend hervorgehoben, um ein klares Verständnis des Ausmaßes und der Schwere eines Vorfalls zu erlangen und um fundierte Entscheidungen über die angemessene Reaktion zu treffen. Inzwischen sind die CERTs der deutschen Bundesländer Teil des Verwaltungs-CERT-Verbunds (VCV), das eine Plattform für den Informationsaustausch in der öffentlichen Verwaltung bietet, einschließlich einer Chat-Plattform und regelmäßiger Treffen, und so eine Institutionalisierung der CERT-Partnerschaften ermöglicht.

Forschungslücke: Bestehende Forschungsarbeiten deuten darauf hin, dass Dashboards ein geeignetes Instrument zur Verbesserung des Cybersituationsbewusstseins sind, aber auch zum Umgang mit

plattformübergreifenden Social-Media-Daten und der Informationsflut, die durch die schiere Menge der über solche Plattformen verfügbaren Daten ausgelöst wird. Obwohl unsere Literaturrecherche eine Vielzahl verschiedener Tools für die Reaktion auf Cybervorfälle ergab, konzentrieren sich diese meist auf ein Netzwerkbewusstsein oder eine bestimmte Datenquelle und bieten somit keinen plattformübergreifenden Überblick, der für das Bedrohungsbewusstsein und das Aufgabenbewusstsein erforderlich ist. Darüber hinaus ergab eine systematische Überprüfung von 54 Ansätzen zur Visualisierung der Cyber-Situationswahrnehmung, dass nur 15 Studien Belege für irgendeine Form von Nutzerevaluierung der vorgeschlagenen Visualisierung nach Abschluss der Entwicklung liefern. Bei der Untersuchung dieser Studien haben wir festgestellt, dass die meisten Evaluierungsstudien nicht auf den CERT-Kontext zugeschnitten sind, nur eine kleine Stichprobe von Teilnehmern verwenden und keine ausführlichen Beschreibungen des Nutzerfeedbacks liefern.

Theoretische Fundierung: Wir folgten dem Designforschungsansatz des Grounded Design, der darauf abzielt, vorläufige Artefakte zu entwerfen, um die in der Praxis begründeten Nutzeranforderungen zu erfüllen und neue Designideen durch die formative Evaluierung von Artefakten im Gebrauch zu generieren. In der HCI-Forschung wurde die Erhebung von Nutzeranforderungen als praxisbezogenes Designwissen charakterisiert, das in hohem Maße präskriptiv und umsetzbar ist, sich aber nur schwer über das Umfeld, in dem es erforscht wurde, hinaus verallgemeinern lässt" Die Durchführung von formativen Evaluationen ist daher geeignet, um Designheuristiken zu identifizieren, die eine Art von praxisinformiertem Designwissen darstellen und darauf abzielen, das Design spezifischer Systeme schrittweise zu verbessern, die in der Regel in Form von Technologieeigenschaften dargestellt werden und ihre Verallgemeinerbarkeit für eine Klasse von Technologien beibehalten. Um sowohl spezialisierte Nutzeranforderungen als auch verallgemeinerbare Designheuristiken zu generieren, befürwortet Grounded Design die Verwendung von Design-Fallstudien als Forschungsrahmen, der idealerweise drei Schlüsselschritte umfasst: (1) Vorstudie, (2) IT-Design und (3) Evaluierung.

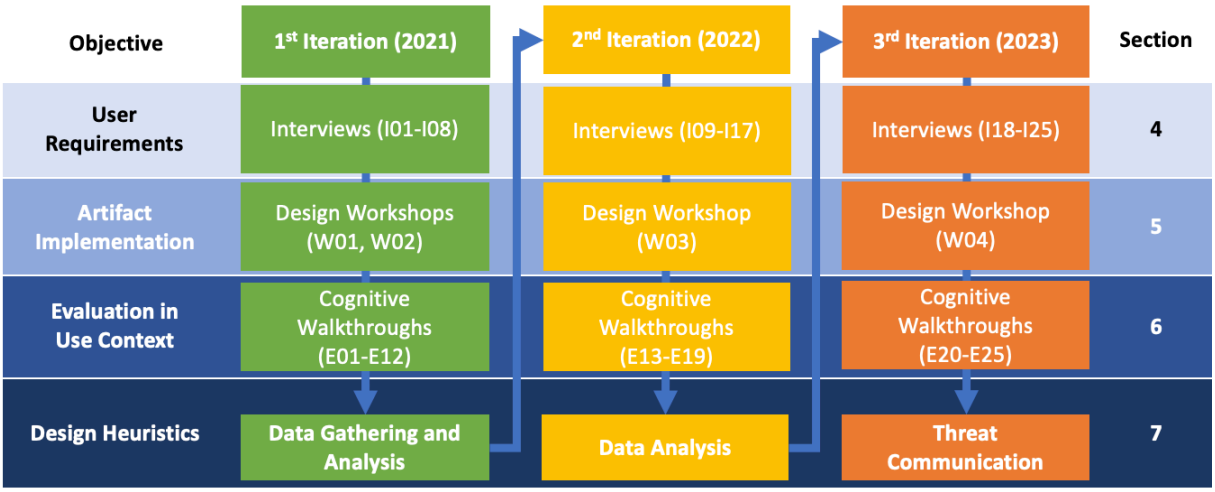

Abbildung 1. Drei Iterationen unserer Design-Fallstudie, bestehend aus Vorstudien zur Ermittlung der Nutzeranforderungen, Design-Interventionen für die Artefakt-Implementierung und Evaluierungssitzungen im Nutzungskontext.

Die empirische Vorstudie soll Einblicke in die sozialen Praktiken geben, bevor eine Intervention erfolgt, und zielt darauf ab, bereits vorhandene Werkzeuge, Medien und deren Nutzung zu beschreiben. In dieser Phase geht es darum, die gegenwärtigen Praktiken unter technologischen, organisatorischen und sozialen Gesichtspunkten zu erfassen, wobei empirische Forschungsmethoden eingesetzt werden, um Probleme, Anforderungen oder Möglichkeiten für die IT-Gestaltung aufzuzeigen. Darauf aufbauend sollte der IT-Designprozess die Beschreibung des innovativen IT-Artefakts sowohl aus der Produkt- als auch aus der Prozessperspektive beinhalten. Durch einen partizipatorischen Ansatz können Entwurfsiterationen auch nach der Einführung der Technologie bei potenziellen künftigen Nutzern fortgesetzt werden. Schließlich sollten Evaluierungsstudien die Analyse Funktionen und Gestaltungsoptionen, die innerhalb des IT-Artefakts realisiert wurden, ermöglichen. Um dauerhafte Veränderungen in den sozialen Praktiken zu erfassen, die sich aus der Einführung von IT-Artefakten ergeben, sind kontinuierliche Dialoge und Reflexionen mit Praktikern erforderlich.

1.1.2 Interne Qualitative Analyse: Interviews mit CERTs (AP1.2)

Die semi-strukturierten Interviews dienten dazu, Einblicke in die Strategien und Technologien der deutschen CERTs zu gewinnen. Der Interview-Leitfaden umfasste Fragen zu mehreren Schlüsselbereichen, darunter (1) die Rollen und Zugehörigkeiten der Befragten, (2) ihre Verfahren zur Meldung von Cyber-Vorfällen, (3) Methoden zur Überwachung von Daten über Cyber-Vorfälle, wie z. B. Indikatoren für eine Kompromittierung, (4) Verfahren zur Analyse, Priorisierung und Validierung der gesammelten Beweise, (5) die Zusammenarbeit zwischen CERTs und (6) die Art und Weise, wie Empfehlungen und Warnungen kommuniziert werden. Nach einer ersten Runde von acht Interviews (n=8, I01-I08), umfasste die zweite Runde neun (n=9, I09-I17) und die dritte Runde weitere acht Interviews (n=7, I18-25). Jedes Interview, das mit dem Einverständnis und der informierten Zustimmung der Teilnehmer durchgeführt wurde, dauerte etwa 60 Minuten. Wir verwendeten eine gezielte Stichprobenstrategie, um in erster Linie Personal auf operativer, aber auch auf taktischer Ebene unter den CERTs der deutschen Bundesländer zu befragen. In den meisten Fällen hatten wir nach unserer anfänglichen E-Mail-Anfrage einen kurzen E-Mail- oder Telefonaustausch mit den jeweiligen CERT-Teamleitern, die uns dann Mitarbeiter auf der operativen Ebene vorschlugen, die für die Durchführung von Interviews im Sinne unserer Forschungsziele geeignet waren. Insgesamt wurden fünfzehn interne Notfallmanager (I01, I04-I07, I14, I17-I24), sechs interne Teamleiter (I02, I08-I10, I12, I16), drei externe Informationssicherheitsbeauftragte (I11, I12, I15) und eine externe Auskunftsstelle für öffentliche Sicherheit (I13) befragt. Insgesamt wurden Teilnehmer aus zwölf Organisationen, darunter neun CERTs und drei andere Organisationen, in unsere Vorstudieninterviews einbezogen.

IKT-Nutzung in deutschen Länder-CERTs: Ein Versuch, die IKT-Nutzung der deutschen staatlichen CERTs zu verallgemeinern, ist in Abbildung 2 dargestellt. Der Technologieeinsatz lässt sich grob in die drei Komponenten der Cyber-Situationserkennung einteilen. Erstens werden Vorfälle entweder von Kunden gemeldet (per Mail oder Telefon) oder von einer Software erkannt (z. B. Intrusion Detection), um ein Netzbewusstsein zu schaffen. Nachdem die ersten Informationen über den Vorfall gesammelt wurden, verwenden die CERTs ein Ticket- und Meldesystem, um ihre Beweise für die Reaktion auf den Vorfall zu sammeln. Zweitens wird das Bewusstsein für Bedrohungen über bewusstseinsorientierte (z. B. Hersteller-Websites, Sicherheitshinweise und Social-Media-Kanäle) und kollaborationsorientierte (z. B. Plattformen für den Austausch von Malware-Informationen, der kollaborative VCV-Chat) Kanäle gesammelt und analysiert. Schließlich werden die gesammelten Informationen verwendet, um einen bestimmten Interessenvertreter mit spezifischen Empfehlungen zu informieren, um (tägliche) Berichte für ausgewählte Interessenvertreter zu erstellen (z. B. einen täglichen Schwachstellenbericht für Ministerien) oder um eine allgemeine Warnung für mehrere Interessenvertreter herauszugeben (wenn große IKT-Infrastrukturen bedroht sind), um ein Missionsbewusstsein zu erreichen.

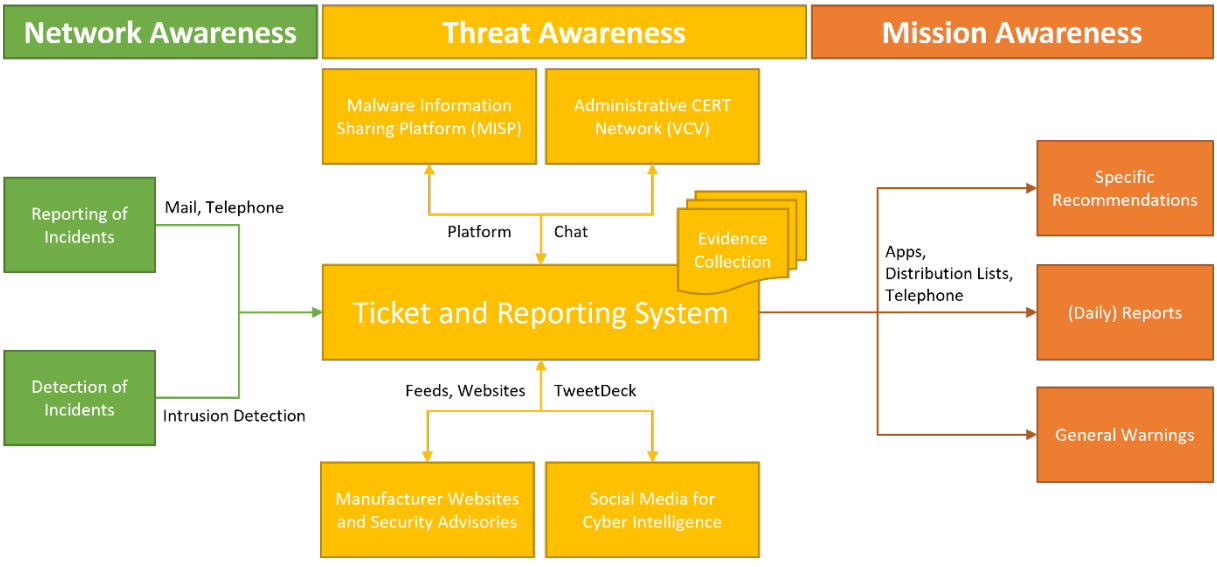

Abbildung 2. Beispiel für eine staatliche CERT-Infrastruktur für Informations- und Kommunikationstechnologie.

Bedrohungsbewusstsein und -verarbeitung: Die CERTs analysieren Hersteller-Websites und Sicherheitshinweise zur Identifizierung von IoCs, doch die variierenden Formate erschweren die Erfassung und erfordern oft individuelle Skripte (I10). Die manuelle Überprüfung der Websites kann täglich bis zu zwei Stunden dauern (I10, I12). Die Datenfülle führt zu Informationsüberlastung, was Priorisierungen nötig macht, und es gibt das Problem redundanter Informationen, das manuell gelöst werden muss (I10). Eine plattformübergreifende Visualisierung könnte nützlich sein, sollte aber modular sein. Einige CERTs überwachen soziale Medien *(*(01, I03, I05, I06*)*, jedoch erfolgt dies oft manuell, wobei Datenschutzvorgaben zu beachten sind *(*I09, I15). Die sich schnell verändernden Bedrohungen erfordern flexible Anpassungen der Technologie (I24).

Missionsbewusstsein und Kollaboration: Zwei CERTs überwachen keine sozialen Medien, sondern erhalten Informationen von anderen staatlichen Stellen oder CERTs (I03). Das VCV dient zur Verifizierung von Informationen, wird aber durch manuelle Prozesse eingeschränkt. IoCs werden über MISP gesammelt und in IDS-Software integriert, was jedoch zu redundanten Einträgen führen kann (I10). Das Wechseln zwischen Tools mindert die Produktivität, daher wäre eine zentralisierte Lösung effizienter (I23). Sechs Teilnehmer sehen Schwierigkeiten in der zielgerichteten Kommunikation von Cyber-Bedrohungen. Anpassbare Warnmeldungen und eine effizientere Genehmigungsplattform für Warnungen werden als notwendig erachtet, um Prozesse zu beschleunigen (I13).

1.1.3 Externe Mixed-Method-Analyse: Repräsentative Bevölkerungsumfragen (AP1.3)

Während der Projektlaufzeit wurden drei repräsentative Bevölkerungsumfragen beauftragt und analysiert. Aufgrund enger Zusammenarbeit sind die Methodik und Ergebnisse der ersten repräsentativen Befragung im Abschlussbericht des Hessischen Ministerium des Innern, für Sicherheit und Heimatschutz (HMdI) hinterlegt. Die Ergebnisse der zweiten und dritten Befragung befinden sich aktuell noch im Publikationsprozess und sind diesem Bericht daher nicht beigelegt.

1.2 Strategien zur Analyse und Kommunikation der Cyberlage (AP2)

Das Ziel von AP2 war die Entwicklung und Verbesserung unterstützender Strategien und Methoden in zwei Iterationen. Auf Basis der erhobenen Anforderungen (AP1) wurden ein Strategiekonzept zur Umsetzung von Open Source Intelligence, Methoden zur Glaubwürdigkeitswertung und Informationspriorisierung sowie Analyse- und Kommunikationsstrategien entwickelt.

1.2.1 Integriertes Methoden- und Strategiekonzept (AP2.1)

Eine systematische und kategorisierte Herangehensweise zur Identifizierung relevanter Informationen in der Cybersicherheit ist essenziell. Die Forschung hebt hervor, dass herkömmliche Sicherheitsmaßnahmen durch gezielte Cyber Threat Intelligence (CTI) ergänzt werden müssen, um umfassenden Schutz zu gewährleisten. Es wurden spezifische Relevanzbegriffe für verschiedene Cybersicherheitstools abgeleitet, um die Informationsüberflutung zu reduzieren und die Relevanz der Daten für unterschiedliche Anwendungsszenarien besser einzuschätzen. Dabei wurden verschiedene Kategorien von Sicherheitsinformationen identifiziert, die zur Verbesserung von Sicherheitsmaßnahmen genutzt werden können, einschließlich Management, Schwachstellenanalyse und Risikobewertung (Abbildung 3).

Category	Context	Precise
Management	Infrastructure	Log files, used hard- and software-information (CPE)
Vulnerability	System weakness	Vulnerability databases, source code, CVSS, CPE, CWE, CAPEC, OVAL
Incident	Infrastructure, network, metadata	Log files, network traffic, payloads, SNORT, YARA
Treatment	Attacker behavior, metadata	Log files, network traffic, IoCs
Intelligence	Blogs, threat information, *vulnerabilities*	IoCs, CVSS, CWE, CVE, attacker behavior
Risk	*Mangement, vulnerability, intelligence*	CPE, CVE, CVSS

Abbildung 3: Kategorien, Kontexte und Kennzahlen relevanter Sicherheitsinformationen

Ein Ziel des Arbeitspakets besteht in der Konzeption eines Frameworks, welche die Erfassung, Vorbereitung, Analyse und Visualisierung von öffentlichen Datenquellen der involvierten CERTs auf operativer und strategischer Ebene unterstützt. Zur Konzeption der ersten Version des Frameworks wurden Literaturquellen und die Ergebnisse der empirischen Vorstudie herangezogen. Im Rahmen der Literaturrecherche wurden insbesondere bestehende Frameworks aus den Bereichen Open Source Intelligence (OSINT), CTI und Social Media Analytics (SMA) analysiert und relevante Komponenten für das Design des Frameworks identifiziert. Die auf Basis der Vorstudie gewonnenen Anforderungen für die Erfassung, Analyse und Kommunikation eines Cyber-Lagebilds wurden genutzt, um das Framework an die Domäne der CERTs anzupassen (Tabelle 2).

(I) Selektion	(II) Erfassung	(III) Verarbeitung	(IV) Analyse	(V) Visualisierung
Vorfallinformationen (Blogs, Soziale Medien, Websites) Vulnerability Reports (NVD) Indicators of Compromise (IoC-Feeds, MISP) Sicherheitshinweise (RSS-Feeds, Websites)	Erfassung (APIs, Feed Parsing, Webscraping) Konvertierung (Austauschformat) Speicherung (Datenbank, Datenformat)	Annotation (ML-Annotation, POS-Tagging) Bereinigung (Duplikate, Stoppworte) Normalisierung (Lemmatisierung, Stammformreduktion)	Aggregation (Cluster, Events, Themen) Extraktion (Entitäten, Metadaten) Klassifikation (Glaubwürdigkeit, Relevanz, Stimmung) Relationen (Netzwerkanalyse)	Filterung (Attribute, Diagramme, Berechnungen, Zusammenfassungen) Auswertung (Erklärbarkeit, Zielorientierung)

Tabelle 2. Framework zur Analyse öffentlicher Daten: Selektion, Sammlung, Verarbeitung, Analyse und Visualisierung

Insgesamt liefert das Konzept ein strukturiertes Framework, das die Erfassung und Nutzung sicherheitsrelevanter Informationen optimiert und damit eine Grundlage für die Implementierung von Strategien und Technologien liefert.

1.2.2 Methoden zur Bewertung der Glaubwürdigkeit und Priorisierung von Inhalten (AP2.2)

Während des Projekts wurden verschiedene KI-gestützte Klassifikationsmethoden zur Bewertung der Glaubwürdigkeit erprobt. Die Ergebnisse zeigen jedoch, dass eine indikatorbasierte Methode einerseits genereller einsetzbar ist und andererseits hinsichtlich der individuellen Anforderungen der CERTs angepasst werden kann. Die entwickelte Komponente analysiert automatisch die Informationen aus sozialen Medien unter Verwendung von Natural Language Processing (NLP), um die Glaubwürdigkeit anhand vordefinierter Nutzer:innen- und Inhalt-basierter Indikatoren zu bewerten. Hinsichtlich einer Information werden die Ergebnisse aggregiert und in die Kategorien „not identifiable“, „not trusted“, „somewhat trusted“ und „trusted“ eingestuft, sodass Expert:innen sofort erkennen können, welchen Inhalten sie vertrauen können. Nutzer:innen haben die Möglichkeit, die Kriterien für die Glaubwürdigkeitsbewertung individuell anzupassen, wie in Abbildung 4 dargestellt.

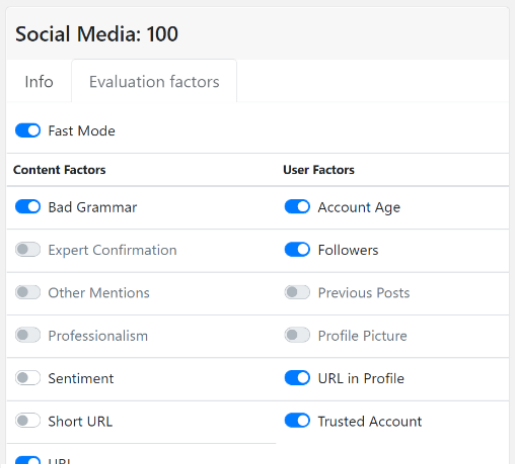

Abbildung 4: Einstellbare Indikatoren der Glaubwürdigkeitskomponente

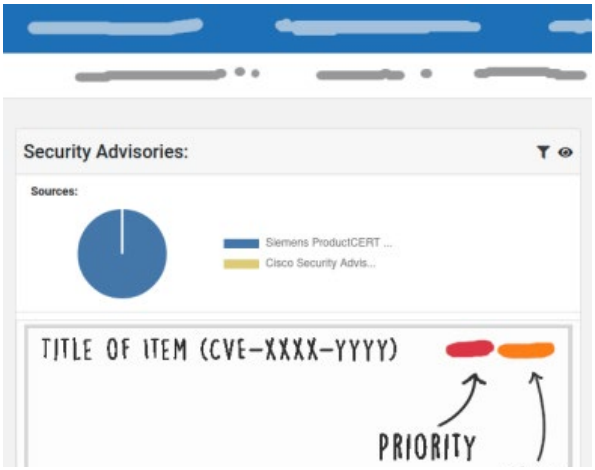

Abbildung 5: Visualisierung der Prioritätsbewertung neben dem CVSS-Severity-Score

Zur Priorisierung wurden verschiedene Techniken zur automatischen Analyse und Filterung von Cyber-Informationen entwickelt. Sie führen zu einer effektiveren Analyse und unterstützen CERTs bei der schnellen Reaktion auf Cyberbedrohungen. Basierend auf Literaturrecherchen und Expert*innen-Interviews wurden verschiedene Prioritätsabhängigkeiten von Schwachstellen identifiziert. Darauf basierend bietet ein Prioritäts-Score eine passgenaue Einschätzung der Relevanz einer Information. Neben der CVSS-Basis-Bewertung wird eine Temporale-Bewertung, welche die Popularität der Information einschätzt und die Existenz von Exploits berücksichtigt, sowie eine Environment-Bewertung, welche die verwendete Soft- und Hardware der CERTs einbezogen. Wie in Abbildung 5 dargestellt, wird der Prioritätsscore im Dashboard neben dem CVSS-Score angezeigt und kann zur Filterung der Inhalte genutzt werden.

1.2.3 Analysestrategie: Maschinelle Analyse des Cyber-Lagebilds (AP2.3)

Die Analysestrategie basierend auf der initialen Version des Hessen3C wurde hinsichtlich der variablen und hoch-individuellen Natur von Cyber-Informationen und deren Relevanz erweitert. Dafür wurde ein Framework entwickelt, welches mit sehr wenigen Beispieldaten Deep Learning-Klassifikatoren trainieren kann, sodass neueste Anforderungen an die Relevanzbestimmung direkt abgebildet werden können. Für das Framework wurden neueste Techniken im Bereich des Deep Learnings entwickelt. Wie in Abbildung 6 zu erkennen ist, wurde dazu zunächst eine großes vortrainiertes Cybersecurity Language Model trainiert, welches als Ausgangslage für alle Klassifikationsaufgaben dienen kann. Folgend wurde ein neues Few-Shot-Learning-Verfahren entwickelt, welches verschiedene Stufen der Feinanpassung vornimmt. Abschließend werden GPT-2 oder GPT-3 genutzt, um neuartige Trainingsdaten zu generieren. Klassifikationsergebnisse mit diesem Framework erreichen nahezu die Qualität eines Klassifikators, der mit der 50-fachen Menge an Trainingsdaten auf herkömmliche Weise trainiert wurde.

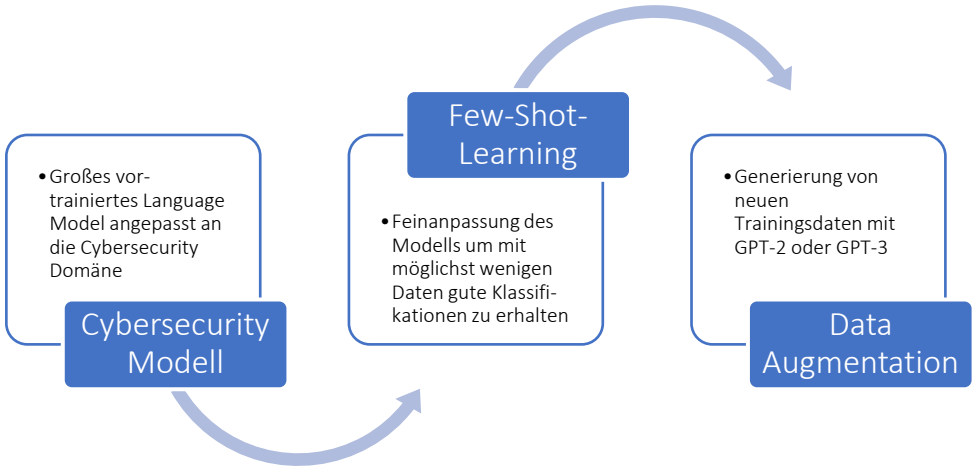

Abbildung 6: Erweiterte und operationalisierte Analysestrategie für die Anwendung von Machine-Learning-Technologien

1.3 Technologien zur Analyse und Kommunikation der Cyberlage (AP3)

Das Ziel von AP3 war die Implementierung eines unterstützenden Demonstrators in zwei Iterationen. Auf Basis der erhobenen Anforderungen (AP1) wurde ein Infrastrukturkonzept entwickelt, um das Konzept und die Implementierung der Basis-, Analyse- und Kommunikationsmodule zu informieren. Auf Basis der entwickelten Analyse- und Kommunikationsstrategien (AP3) sowie benutzerzentrierter Evaluationen (AP4) werden das Konzept sowie die Module in einer zweiten Iteration verbessert.

1.3.1 Integriertes Funktions- und Technologiekonzept (AP3.1)

Das Infrastrukturkonzept ist in Abbildung 2 visualisiert und gliedert sich in unterschiedliche Basis-, Analyse- und Kommunikationsmodule. Das **Basismodul** operiert vollständig im Backend und umfasst die Task-, Scraping- und Interface-Komponenten. Die Task Component nimmt zunächst die

Suchaufträge der Benutzer*innen entgegen. In diesen Suchaufträgen sind neben Keywords und weiteren Metadaten insbesondere die Datenquellen definiert, aus denen Informationen bezogen werden. Für Datenquellen, die nicht über eine API angebunden sind (z.B. Advisories und Websites), liefert die Scraping Component mitsamt Web-Scraping-Technologien die gefundenen Ergebnisse an die Task Component zurück; sind die Datenquellen jedoch über eine API erreichbar (z.B. Social Media, Vulnerability Databases), liefert die Interface Component die angeforderten Informationen an die Task Component zurück. Bevor die Task Component die Ergebnisse in der Datenbank (Remote Storage) speichert, wird jedoch die Analysekomponente zur Anreicherung der Daten angesprochen.

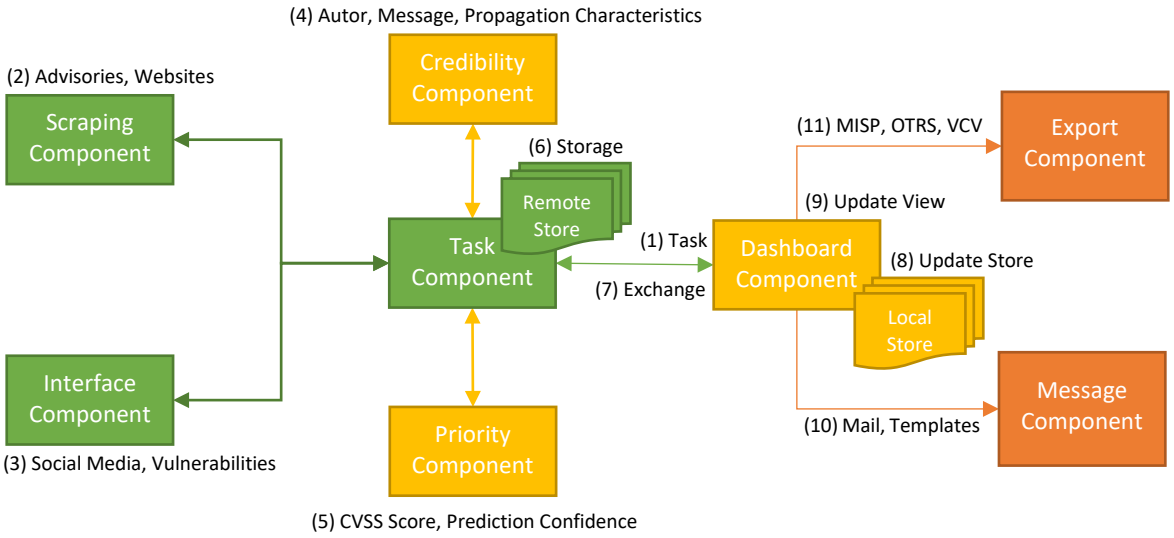

Abbildung 7. Aktuelle Architektur des CYWARN-Demonstrators mit Basiskomponenten (grün), Analysekomponenten (gelb) und Kommunikationskomponenten (orange)

Das **Analysemodul** umfasst die Credibility-, Priority- (je im Backend operierend) und Dashboard-Komponenten (im Frontend operierend). Die Task Component schickt die gesammelten Daten zunächst an die Credibility Component, um die Glaubwürdigkeit der Daten zu berechnen und anschließend an die Priority Component, um die Daten nach Relevanz zu priorisieren. Nachdem die analytisch angereicherten Daten in der Datenbank gesichert wurden, werden diese der Task Component an die Dashboard Component geschickt. Die neuen Daten werden nun in der Dashboard Component visualisiert und in einer lokalen Datenbank (Local Store) vorgehalten, damit Benutzer*innen analytische bestimmte analytische Operationen (z.B. interaktive Filterung der Daten) direkt im Frontend vornehmen können.

An das Dashboard ist das **Kommunikationsmodul** angebunden, welches die Message- und Export-Komponenten anbindet. Die Message Component erlaubt es, Informations- und Warnmeldungen über verschiedene Kanäle (z.B. E-Mail) an Zielgruppenakteure zu schicken. Die Formulierung der Meldungen wird dabei durch Vorlagen (Templates) unterstützt. Darüber erlaubt es die Export Component, relevante Informationen in andere Anwendungen (z.B. Excel, MISP, OTRS, VCV) zu exportieren. Die Module des Infrastrukturkonzepts werden im Folgenden jeweils detaillierter mitsamt den zugehörigen, empirisch erhobenen Anforderungen vorgestellt.

1.3.2 Implementierung des Cyber Threat Observatory (AP3.2, AP3.3)

Das Frontend unseres Artefakts ist eine Webanwendung auf Basis von Vue.js, inklusive Bootstrap für responsives Design und Chart.js für die Datenvisualisierung. Neben einigen lokalen Filtermöglichkeiten werden alle anderen Aktionen des Frontends, wie die Suche nach Beiträgen in offenen und sozialen Medien oder die Verwaltung von Nutzern, an ein API-Backend weitergeleitet. Das Backend ist als Service realisiert und folgt dem Paradigma einer webbasierten und serviceorientierten Architektur, die eine Datenbereitstellung für externe Organisationssoftware ermöglicht. Es handelt sich um eine Python-Anwendung, die das Flask-Framework für RESTful-Webdienste und die Datenbank MongoDB

für die dokumentenorientierte Datenverwaltung nutzt. Mehrere Bibliotheken erleichtern die automatisierte und kontinuierliche Echtzeit-Sammlung von Daten aus offenen Quellen wie NVD-Schwachstellen, IoCs und RSS-Feeds oder Social-Media-Quell-APIs, darunter Flickr, Reddit, Tumblr, Twitter und YouTube. Die schnittstellenbasierte Programmierung wurde angewandt, um eine modulare Anwendung zu schaffen, die die Erweiterung dieser implementierten Quellen in zukünftigen Iterationen ermöglicht. Um die Vielfalt der Datenzugänglichkeit und -strukturen zu überwinden, werden alle Entitäten gemäß der ActivityStreams 2.0 Core Syntax verarbeitet und gespeichert. Für die persistent gespeicherten Daten hat der Systembetreiber die Möglichkeit, ein Verfallsdatum festzulegen, an dem die Daten gelöscht werden. Außerdem wurden eindeutige Datenbankkennungen verwendet, um zu verhindern, dass gesammelte Daten gespeichert werden, die bereits in der Datenbank vorhanden sind und daher redundante Informationen darstellen würden.

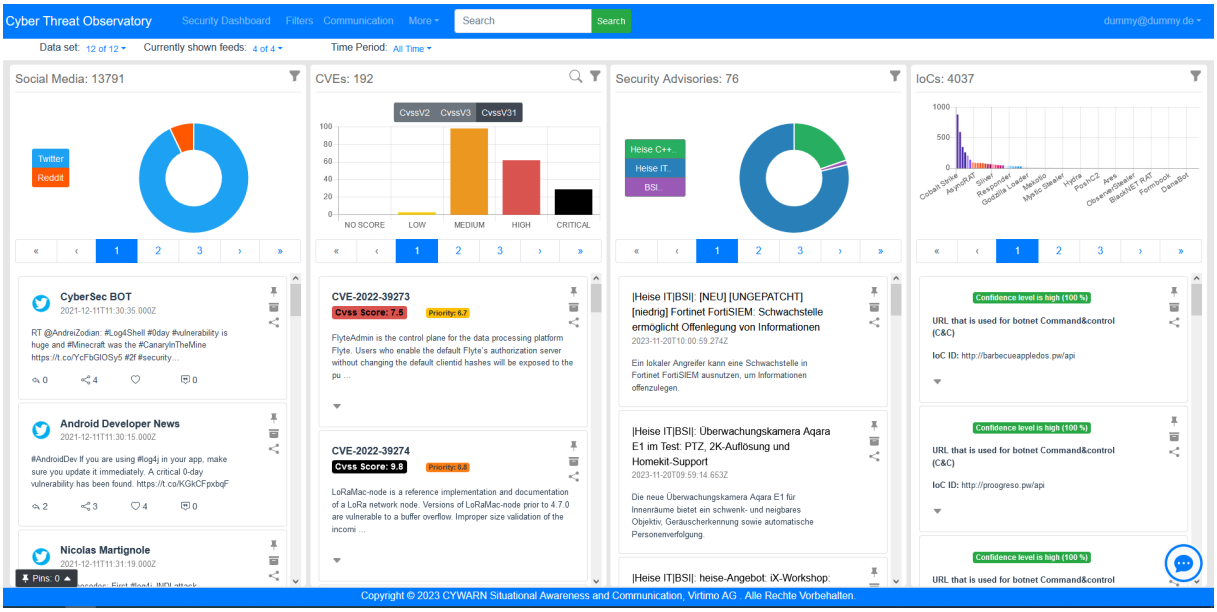

Abbildung 8. Analysekomponente: Dashboard des Cyber Threat Observatory

Die in Abbildung 5 dargestellte Schnittstelle umfasst bis zu vier Feeds mit Sicherheitshinweisen, CVEs, IoCs und Daten aus sozialen Medien. Die Sicherheitshinweise werden über RSS-Feeds eingebettet, die von Software- und Hardwareanbietern bereitgestellt werden, und die API der NVD-Datenbank wird verwendet, um den CVE-Feed mit dokumentierten Schwachstellen zu füllen. Darüber hinaus haben wir uns entschieden, die ThreatFox-Plattform für den Empfang von IoCs und individuelle Plattform-APIs zu nutzen, um Informationen aus sozialen Medien (z. B. Reddit oder Twitter) zu sammeln. Für jeden Feed werden spezifische Diagramme und ein unterschiedlicher Satz verfügbarer und charakteristischer Informationen pro Eintrag angezeigt (z. B. eine textliche Beschreibung und der CVSS-Score für CVEs; der Autor, der Textkörper und die Retweets eines Twitter-Posts). Der angezeigte Datensatz kann in der oberen linken Ecke ausgewählt werden und basiert auf vordefinierten individuellen Parametereinstellungen, die zur Abfrage der verschiedenen Quellen verwendet werden. Auf Wunsch können Nutzer einzelne oder alle Feed-Einträge ein- oder ausblenden und wichtige Einträge anpinnen, die dann hervorgehoben und im schwarzen unteren Pin-Menü für den schnellen Zugriff angezeigt werden. Um die angezeigten Informationen zu filtern, können die Nutzer entweder auf die interaktiven Diagramme klicken (z. B. auf den kritischen Balken, um nur kritische Sicherheitslücken innerhalb des CVE-Feeds anzuzeigen) oder ein erweitertes Filtermenü mit zusätzlichen Optionen öffnen (z. B. auch Filterung nach der CVE-ID oder dem betroffenen Produkt einer Sicherheitslücke). In der oberen rechten Ecke ermöglicht ein Volltextsuchfeld die gleichzeitige Suche nach Schlüsselwörtern in den verschiedenen Feeds, während ein Algorithmus zur Berechnung der Priorität von CVE-Einträgen anhand von Kontextdaten verwendet wird. Um die Sammlung und Verarbeitung von Daten in der Anwendung transparent zu machen, enthält jede angezeigte Information einen Link zu einer externen Quelle (z.B. die NVD für die Schwachstellen), um die Überprüfungsmöglichkeiten zu verbessern.

Darüber hinaus haben wir in der rechten unteren Ecke eine Mastodon-Chat-Schnittstelle implementiert, damit CERT-Mitglieder ihr Wissen über aktuelle Cybervorfälle oder wichtige Schwachstellen austauschen können. Um erkannte Bedrohungen mit Dritten zu teilen, wurde das Kommunikationsmodul implementiert (Abbildung 6). Das Modul unterstützt die Erstellung von Warnmeldungen anhand einer Vorlage, um die Informationen und die Sprache an die jeweilige Zielgruppe anzupassen. Mit Hilfe von Dropdown-Menüs kann der Benutzer Informationen zur Meldung (z.B. Art des Vorfalls, betroffene Zielgruppe, Hyperlinks und zugehörige CVEs) und zur betroffenen Software (z.B. Name, Version, Betriebssystem und Art des Angriffs) auswählen sowie eine Risikoklassifizierung und Gegenmaßnahmen angeben. Abhängig von den ausgewählten Eigenschaften enthält der generierte Text Platzhalter, um die Merkmale der Bedrohung und die Anweisungen für Gegenmaßnahmen weiter zu spezifizieren. Während es auch möglich ist, die Sensibilität von Informationen auf der Grundlage des Ampelprotokolls (TLP) zu klassifizieren, war es uns aufgrund der begrenzten Zeit des Forschungsprojekts nicht möglich, ein Genehmigungssystem mit ausreichender Reife für die Evaluierung mit CERT-Mitarbeitern zu implementieren. Es stehen jedoch verschiedene Kanäle (derzeit E-Mail, Reddit und Twitter) zur Verfügung, um die zuvor generierten Warnungen zu verbreiten und die betroffenen Akteure zu erreichen.

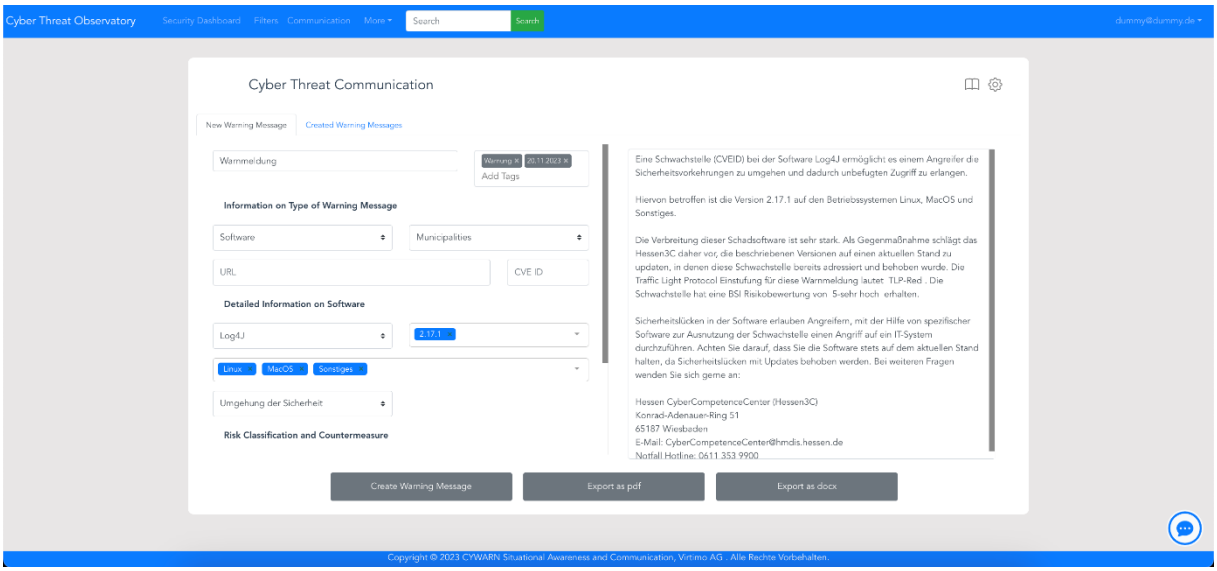

Abbildung 9. Kommunikationskomponente: Interaktive und formularbasierte Erstellung von Warnmeldungen

1.4 Evaluations- und Verstetigungsstrategie (AP4)

Das Ziel von AP4 war die Evaluation der entwickelten Methoden und Strategien sowie des Demonstrators durch die Anwendungs- und assoziierte Partner, sodass diese fortlaufend verbessert werden können (AP2, AP3). Des Weiteren wurden Konzepte zum Transfer der Projektergebnisse auf andere Anwendungsbereiche im Sinne der wirtschaftlichen Verwertung sowie zur Verstetigung der Projektergebnisse bei den Anwendungspartnern nach Projektende erarbeitet.

1.4.1 Evaluation der konzipierten Strategien und Technologien (AP4.1, AP4.2)

Unser Evaluierungsansatz umfasste kognitive Walkthroughs mit Nutzern, gefolgt von kurzen halbstrukturierten Interviews. Für die kognitiven Walkthroughs definierten wir zunächst drei Aufgaben: die Suche nach spezifischen Informationen über verdächtige Aktivitäten von drei IP-Adressen, die Bewertung kritischer Schwachstellen einer bestimmten Browserversion und die Beantwortung einer Chat-Anfrage zu SSL-Zertifikatsproblemen in einer Konferenzsoftware. Die Teilnehmer wurden angewiesen, ihre Gedanken zu äußern, während sie durch die Aufgaben navigierten, wobei sie dem Think-Aloud-Protokoll folgten. Während unserer ersten Evaluationssitzungen stellten wir fest, dass die Teilnehmer es vorzogen, das Tool frei zu erkunden oder ihre eigenen, aus der täglichen Praxis bekannten Aufgaben zu wählen. Daher boten wir den

nachfolgenden Teilnehmern die Möglichkeit, die vordefinierten Aufgaben (N=3) oder eine freie Erkundung des Tools (N=22) durchzuführen, möglicherweise mit eigenen Aufgaben im Kopf. Die anschließenden halbstrukturierten Interviews waren so konzipiert, dass sie die Teilnehmer zur Reflexion des Bewertungsprozesses aufforderten und somit einen Vergleich der Teilnehmeraussagen in Bezug auf die Kernthemen der Bewertung ermöglichten. Die Interviewfragen konzentrierten sich auf die wahrgenommene (1) Benutzerfreundlichkeit und visuelle Präsentation, (2) Datenquellen und Informationsarten, (3) Such- und Filterfunktionalität, (4) Informationspriorisierung und Glaubwürdigkeitsbewertung sowie (5) Zusammenarbeit und Kommunikationsfunktionalität. Die Evaluation mit deutschen CERT-Mitarbeitern wurde in drei Iterationen durchgeführt, mit zwölf Teilnehmern in der ersten Runde (N=12, E01-E12), sieben in der zweiten Runde (N=7, E13-I19) und sieben in der dritten Runde (N=7, E20-25). Jede Bewertungssitzung, die nach Einholung der informierten Zustimmung über Web-Konferenz-Tools durchgeführt wurde, dauerte etwa 60 Minuten. Der Teilnehmerkreis bestand aus neunzehn internen Incident Managern (E05-E10, E13-19, E20-E25), drei internen Teamleitern (E02, E11, E12) und drei auf HCI, Informationssicherheit und künstliche Intelligenz spezialisierten Forschern (E01, E03, E04). Alle Interviews wurden aufgezeichnet und mit dem Einverständnis der Teilnehmer transkribiert.

Präsentation und Benutzerfreundlichkeit des Tools: Die meisten Teilnehmer bewerteten das Dashboard positiv und schätzten die Übersichtlichkeit und Benutzerfreundlichkeit (E22). Die Integration mehrerer Datenquellen wurde als effizient angesehen, obwohl Bedenken über eine mögliche Überforderung durch zu viele Informationen geäußert wurden (E02, E03, E06). Interaktive Diagramme wurden als hilfreich empfunden, einige Teilnehmer wünschten jedoch andere Visualisierungen, z. B. zur Häufigkeit von Stichworten (E04, E08, E09, E10). Vorschläge wie das Einklappen von Feed-Einträgen (E03, E07, E08) und die Optimierung des Layouts durch Ausblenden weniger relevanter Feeds wurden gemacht (E01, E02, E04, E07, E10).

Datenquellen und Arten von Informationen: Teilnehmer begrüßten die Datenfeeds, hatten aber unterschiedliche Anforderungen. Einige wünschten eine Vorfilterung der CVEs auf relevante Produkte (E05, E10) oder Informationen über neue/aktualisierte CVEs (E07). Während einige Teilnehmer IoCs als weniger relevant empfanden (E05, E06), sahen andere deren Integration als nützlich an (E02, E08, E09). Es wurden Bedenken hinsichtlich der Vollständigkeit des Sicherheitshinweis-Feeds geäußert (E07, E10). Der Social-Media-Feed wurde unterschiedlich bewertet, aber die Verbindung zwischen verschiedenen Feeds und die Integration weiterer Datenquellen wurden allgemein als wichtig angesehen (E01, E03, E05, E06, E07, E08, E10).

Feedübergreifende Such- und Filterfunktionalität: Die globale Suche wurde positiv bewertet, obwohl es manchmal schwierig war, Suchfilter zurückzusetzen (E02, E12). Vorschläge beinhalteten die Verwendung von Regex (E03, E12) und mehr Funktionen ähnlich der Unix-Befehlszeile. Filteroptionen wurden als nützlich empfunden, aber es gab Vorschläge zur Verbesserung, wie z. B. das Filtern nach Herstellern/Produkten im CVE-Feed (E01, E04, E11) und die Integration von Infrastrukturmodellen (E11). Die Möglichkeit, durch Diagramme zu filtern, wurde schnell entdeckt und als intuitiv angesehen (E01, E03).

Glaubwürdigkeit und Prioritätsbewertung: Einige Teilnehmer sahen die Such- und Filterfunktion als hilfreich für die Priorisierung (E03, E07), wiesen jedoch darauf hin, dass der CVSS-Score oft als Obergrenze gesehen wird (E14, E15, E16). Umweltkennzahlen wurden als wichtige Ergänzung zur Bewertung der Relevanz angesehen (E13, E14, E15, E16). Die Teilnehmer schlugen vor, Cluster- und Deduplizierungsalgorithmen zu verwenden, um die Informationsflut zu reduzieren (E05, E13, E16). Die Glaubwürdigkeit der Informationen sollte durch vertrauenswürdige Autoren und Quellen gewährleistet werden, insbesondere in sozialen Medien (E01, E05, E06, E17, E18, E19).

Kollaboration und Kommunikation: Einige Teilnehmer empfanden den kollaborativen Chat als ablenkend, während andere den Nutzen in der Kommunikation innerhalb und außerhalb der Organisation sahen (E02, E07, E11). Die Integration von Fallverwaltungsfunktionen und halbautomatische Nachrichtenvorlagen wurde geschätzt (E23, E24). Vorschläge beinhalteten die Auswahl mehrerer Zielgruppen für Warnmeldungen (E25) und eine erweiterte Exportfunktion für Datensätze (E01, E02, E03, E04, E07, E08, E09).

1.4.2 Integriertes Transfer- und Verstetigungskonzept (AP4.3, AP4.4)

Das Cyber Threat Observatory zielt darauf ab, die drei Einsatzszenarien (Use Cases) der Erfassung, Analyse und Kommunikation von Cyberbedrohungen und Sicherheitslücken zu unterstützen. Obwohl das CTO diese Szenarien bereits gut abbildet, haben sich neue Forschungsbedarfe zur Projektlaufzeit ergeben, die notwendigerweise adressiert werden müssen, um eine Nutzung in vielseitigen Einsatzumgebungen zu ermöglichen.

Problemdarstellung und Lösungsansatz: Im Hinblick auf die **Erfassungskomponente** wurde deutlich, dass die *Anbindung öffentlicher Datenquellen* mit einem erheblichen Wartungsaufwand verbunden sein kann. Durch einerseits Veränderungen oder dem Wegfall von Programmierungsschnittstellen (APIs), wie zuletzt bei Twitter bzw. X der Fall, und andererseits Änderungen an der Struktur von webbasierten Security Advisories, die mangels einer API mitunter per Webscraping erfasst werden, muss deren Anbindung regelmäßig gewartet werden. Daher muss erneut evaluiert werden, welche sozialen Netzwerke primär eingebunden werden sollten. Weiterhin forciert das BSI die Implementierung und Verbreitung des Common Security Advisory Framework (CSAF), womit Security Advisories in Zukunft vermehrt über ein standardisiertes Format bereitgestellt werden, welches in das CTO eingebunden werden muss. Ein wichtiges Kriterium zur *Vorfilterung der Informationen* die aus öffentlichen Datenquellen bezogen werden liegt in der Soft- und Hardware, welche in der betreuten Organisation eingesetzt wird und daher besonders relevant für die Reaktion auf Cyberbedrohungen ist. Mangels einer Integration mit organisationaler Software müssen die genutzten Produkte im CTO derzeit noch manuell gepflegt werden, sodass die Anbindung einer Produktdatenbank vorgenommen werden muss, um Informationen über genutzte Produkte automatisch bereitzustellen.

Im Rahmen der **Analysekomponente** wurde deutlich, dass CERT-Mitarbeiter*innen einerseits selektiv anderen Cybersecurity-Expert*innen zum Informationsaustausch in sozialen Medien folgen, andererseits aber auch über Suchanfragen nach Informationen suchen, welche in sozialen Medien häufig schneller als in offiziellen Kanälen veröffentlicht werden. Durch die schiere Menge an Informationen aus sozialen Medien kann es jedoch zu einem Informationsüberfluss kommen, der von einzelnen Mitarbeiter*innen nicht bewältigt werden kann. In den Jahren nach Bewilligung des CYWARN-Projekts gab es in der KI einen Boom durch den Einsatz großer Sprachmodelle (Large Language Models, kurz: LLMs) wie z.B. BERT oder GPT-3, welche die Erkennung relevanter Cybersicherheitsinformationen mit einer hohen Genauigkeit unterstützen können. In Zukunft wird daher der rechtskonforme Einsatz von *LLMs zur Filterung von Informationen* aus sozialen Medien forciert und mit Ansätzen der erklärbaren KI kombiniert, um den Entscheidungsprozess des Algorithmus für Entscheidungsträger*innen transparent zu visualisieren. Im Rahmen der iterativen Evaluation des CTO wurde ebenso deutlich, dass die *Analyse von Bezügen zwischen den unterschiedlichen Datenströmen (Feeds)* einen besonderen Mehrwert schaffen kann, bspw. wenn bei der Auswahl eines Security Advisories automatisch zugehörige Schwachstellen, IoCs, Social-Media-Posts und betroffene Produkte hervorgehoben werden.

Die aktuelle Fassung der **Kommunikationskomponente** bietet noch keine Unterstützung des Berichtswesens in CERTs. Durch die manuelle, mehrstündige Bearbeitung von Security Advisories für werktägliche Schwachstellenberichte besteht ein besonderes Potenzial in der *(semi-)automatisierten Generierung von Schwachstellenberichten*, die insbesondere von einer Integration des CSAF-Standards und einer automatisierten Produktfilterung profitieren. Obwohl in CYWARN bereits ein exemplarischer, plattformübergreifender Formulierungs- und Freigabeprozess von Warnmeldungen implementiert wurde, unterscheiden sich die organisationalen Anforderungen mitunter deutlich. Es ist daher notwendig, das Modul mit internationalen oder privatwirtschaftlichen Stakeholdern abzugleichen und die Funktionalität entsprechend zu erweitern. Dies inkludiert beispielsweise die Erweiterung der vorhandenen Nachrichtenvorlagen einschließlich ihres variablen Detailgrades für bestimmte Empfänger, eine Unterstützung adäquater Kommunikationskanäle, die Möglichkeit, Anhänge mit Detailinformationen zu verschicken und die Option, für jede Nachricht Inhalt, Ersteller, Absender und Empfängergruppen für eine rechtskonforme Archivierung nachzuverfolgen.

Szenario	Problemdarstellung	Lösungsansatz
(I) Erfassung	Wartungsaufwand durch öffentliche und nicht-standardisierte Datenquellen	Aktualisierung der Quellenanbindung und Integration des CSAF-Standards
	Manuelle Produktfilterung durch fehlende organisationale Integration	Einbindung des CTO an organisationale Produktdatenbanken
(II) Analyse	Keine transparenten Algorithmen zur Relevanzbewertung in sozialen Medien	Nutzung von rechtskonformen LLMs mit erklärbarer KI
	Isolierte Analyse- und Filterfunktionalitäten pro Datenquelle	Relationsanalyse zwischen unterschiedlichen Datenquellen
(III) Kommunikation	Fehlende Unterstützung des werktäglichen Berichtswesens	Automatische Generierung werktäglicher Schwachstellenberichte
	Anpassung des Formulierungs- und Freigabeprozesses von Warnmeldungen	Integration weiterer Templates und Warnkanäle

Tabelle 3. Problemstellungen und Lösungsansätze in Bezug auf die Use Cases des Cyber Threat Observatory

Marktpotenzial: In CYWARN wurde eine Marktanalyse über bestehende Cybersecurity-Anwendungen durchgeführt, die zu einer Anwendungstypologie führte, die sich an zentralen Szenarien der Cybersicherheitspraxis orientieren. Das CTO lässt sich innerhalb dieser Typologie einerseits der Klasse *Intelligence* zuordnen, da es sowohl IoCs als auch andere Arten von CTI aus unterschiedlichen Quellen bereitstellt. Es umfasst dabei Funktionalitäten von fortgeschrittenen Open Source und Social Media Intelligence (SOCMINT) Tools, da es bedrohungsrelevante Informationen aus sozialen Medien und aus anderen öffentlich Quellen automatisiert abruft, analysiert, und in Feeds bündelt. Andererseits lässt es sich auch innerhalb der Klasse *Vulnerability* verorten, da es mit Security Advisories (SecAdv) und Sicherheitslücken (CVEs) aus der National Vulnerability Database (NVD) [20] Informationen bündelt und Funktionalitäten zu deren Analyse bietet. Damit deckt das CTO gleichzeitig Funktionalitäten ab, die zum Portfolio von Vulnerability Databases oder Vulnerability Management Tools gehören.

Tools	IOCs	CVEs	SecAdv	SOCMINT	IRM	Relevanz	Glaubwürdigkeit	CSCW	CERT-Zuschnitt
Pulsedive	X	-	-	X	-	X	-	X	-
ThreatConnect TIP	X	X	-	-	-	X	-	X	-
Anomali Threatstream	X	X	-	-	-	X	X	X	-
SOCRadar XTI	X	X	-	X	-	X	-	X	-
Cyware Intel Exchange	X	X	X	-	-	X	X	X	-
ThreatQuotient THREATQ	X	X	X	-	-	X	-	X	-
Vulners	X	X	X	-	-	X	-	-	-
TheHive	X	X	-	-	-	-	-	X	X
Threatfox*	X	-	-	-	-	-	-	-	-
IntelOwl*	X	X	-	-	-	-	-	X	-
MISP*	X	X	-	-	-	-	-	X	X
Cyber Threat Observatory	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Tabelle 4. Vergleich der Eigenschaften des entwickelten Cyber Threat Observatory mit marktverfügbaren, kommerziellen und nicht-kommerziellen (*) Cybersecurity-Tools mit ähnlichen Funktionalitäten.

Während mehrere in der Forschung entwickelte Technologieartefakte die Aggregation und Analyse externer Bedrohungsinformationen ermöglichen, sind diese nur in Ausnahmefällen als einsatzbereite Tools verfügbar. Ein Blick auf marktverfügbare Tools (Tabelle 2) mit ähnlichen Funktionalitäten verdeutlicht die Besonderheiten des CTO. Das CTO bietet eine einzigartige Kombination von individuell anpassbaren IoC, CVE, Security Advisory und SOCMINT Feeds, während vergleichbare Tools nur einen Teil dieser Informationen liefern. Insbesondere eine Integration von SOCMINT wird selten angeboten, obwohl dies einen Mehrwehrt bietet, da Sicherheitsforscher*innen neuartige Bedrohungen häufig auf Social Media Plattformen diskutieren. Außerdem werden Security Advisories häufig nur aus ausgewählten Quellen und Formaten bezogen. Das CTO sieht hingegen eine Integration von Security Advisories aus einer Vielzahl von Quellen in diversen Formaten vor. Weiterhin nimmt das CTO eine algorithmische Relevanz- oder Prioritätsbewertung von strukturierten Bedrohungsinformationen wie CVEs vor, die künftig durch Schnittstellen zu IRM-Tools auch die automatisierte Informationsfilterung

nach Relevanz für die betreute IT-Infrastruktur erlauben soll. Darüber hinaus enthält das CTO eine Komponente zur Glaubwürdigkeitsbewertung von unstrukturierten SOCMINT-Daten (z.B. von Reddit, Twitter/X), welche ein Alleinstellungsmerkmal darstellt. Das CTO erlaubt weiterhin nicht nur einen Austausch zu Bedrohungsinformationen durch kollaborative Funktionalitäten wie einen integrierten Chat, sondern ermöglicht auch die teilautomatisierte Erstellung von zielgruppengerechten Warnmeldungen und Reports. Abschließend ist das CTO durch seinen nutzerzentrierten Designansatz direkt auf die CERT-Domäne zugeschnitten.

Folglich besetzt das CTO, gerade mit Blick auf die Zielgruppe der CERTs und ITSBs, eine Nische, die für VIRTIMO ein vielversprechendes Marktpotential aufweist. Die skizzierten Transfer- und Verwertungskonzepte sind darauf ausgelegt VIRTIMO zu befähigen, auf Basis der entwickelten Strategien und Technologien ein Beratungs- und Umsetzungskonzept als Dienstleistung für sicherheitsrelevante Akteure bzw. Kunden im Folgejahr des Projekts aufzubauen. Daher können Organisationen und Unternehmen bereits kurzfristig von Architekturentwicklungen („Cloud“ für Dienstleitungsprozesse mit „Pay-per-Use“-Modellen) im Kontext von Vollserviceangeboten profitieren, indem die Lösungen in das „Virtimo Publication Framework“ integriert werden. Während Landes-CERTs als Beratungsstelle für KMU dienen, ist davon auszugehen, dass größere Unternehmen mit mehr als 250 Mitarbeiter*innen ein Interesse am eigenständigen Betrieb einer Lösung, wie sie in CYWARN als Demonstrator erprobt wurde, aufweisen. Sollte VIRTIMO bei ca. 15.000 großen Unternehmen in Deutschland eine Markterschließung von vier Prozent in fünf Jahren nach Überführung des Demonstrators in ein Produkt erreichen, ergäbe sich bei 150 Euro Lizenzgebühren pro Monat im betrachteten Zeitraum ein Umsatz über dreieinhalb Millionen Euro. Gleichzeitig stellen produktbegleitende Dienstleistungen wie Schulungen, Wartungen und Inbetriebnahme eine weitere Verwertungsmöglichkeit dar, die auf zehn Prozent der Lizenzkosten des Produktes kalkuliert wird.

1.5 Ethische, soziale und rechtliche Technikfolgenabschätzung (AP5)

Das Ziel von AP5 war die Abschätzung der ethischen, sozialen und rechtlichen Auswirkungen der entwickelten Strategien und Technologien im Hinblick auf die Analyse und die Bewertung der Cyberlage sowie die Kommunikation von Cyberbedrohungen. Dies inkludierte insbesondere den Datenschutz und die Privatsphäre beteiligter und betroffener Akteure. Nach einer initialen Erfassung wurden die Ergebnisse in zwei Iterationen evaluiert und verfeinert.

1.5.1 Ethisch-soziale Implikationen und Technikfolgenabschätzung (AP5.1)

In diesem Teilarbeitspaket ist die Durchführung einer iterativen, ethischen und sozialen Technikfolgenabschätzung in allen Projektbereichen (AP2, AP3) und die Erarbeitung von Design- und Entwicklungsimplikationen mithilfe der Methode des Value-Sensitive Designs (VSD) vorgesehen. In dem Rahmen wurden sowohl eine systematische Literaturrecherche durchgeführt und die durchgeführten Interviews (AP1.2) nachanalysiert. Im Rahmen dieser Studie wollten wir Erkenntnisse über die Einsatzszenarien, die technischen Funktionalitäten, die algorithmischen Ansätze, die Datenquellen und die möglichen ethischen, rechtlichen und sozialen Implikationen der Technologien gewinnen. Da privatwirtschaftliche Forschung oft nicht öffentlich verfügbar ist, wurden nur akademische Forschungsarbeiten untersucht. Es wurden vier Literaturdatenbanken durchsucht. Der Umfang des Reviews beschränkte sich auf englischsprachige Publikationen in begutachteten Zeitschriften und Tagungsbänden, die vom Beginn der Abdeckung der jeweiligen Datenbanken bis Ende Mai 2021 veröffentlicht wurden. Die Volltextsuche erfolgte mit folgendem Suchausdruck: *("cyber security" OR cybersecurity OR "information security" OR cybercrime) AND (OSINT OR SOCMINT OR WEBINT OR "open-source intelligence" OR "social media intelligence" OR "web intelligence")*. Anwendung vordefinierter Ausschlusskriterien wurden 73 Publikationen anhand eines Kategoriensystems mithilfe von Excel quantitativ ausgewertet.

Um relevante Werte und Wertekonflikte zu identifizieren, führten wir eine Fokusgruppenworkshop innerhalb des Forschungsteams (N=7) durch und analysierten die in der zweiten Runde durchgeführten semistrukturierten Experteninterviews (N=9). Die Fokusgruppe umfasste sieben Teilnehmende, die

Teil des CYWARN-Forschungsteams sind. Nach einem Input über VSD und ein hypothetisches Nutzungsszenario von OSINT-Technologien aus der CYWARN-Forschung wurden die Teilnehmenden gebeten, ein Brainstorming zu ethischen, rechtlichen und sozialen Problemstellungen durchzuführen und diese anschließend mit Blick auf potenziell betroffene Werte und Wertekonflikte zu diskutieren. Die semi-strukturierten und online durchgeführten Interviews beinhalteten ebenfalls Fragen, die darauf abzielten für Stakeholder relevante Werte und Wertekonflikte herauszuarbeiten. An den neun Interviews nahmen Personen aus drei Stakeholder-Gruppen (CERT-Mitarbeitende, potenziell OSINT-Nutzende, zivilgesellschaftliche Expert*innen) teil. Transkripte der Audiomittschnitte wurden im Rahmen einer qualitativen Inhaltsanalyse mit der Software MAXQDA ausgewertet. Die deduktiv auf Basis von verwandter Literatur und induktiv während der Analyse entwickelten Kategorien wurden in einem Codebuch mit ergänzenden Kodier-Regeln und Beispielen definiert. Es wurden zehn Kategorien zu unterschiedlichen Werten sowie eine Kategorie zu Wertekonflikten verwendet.

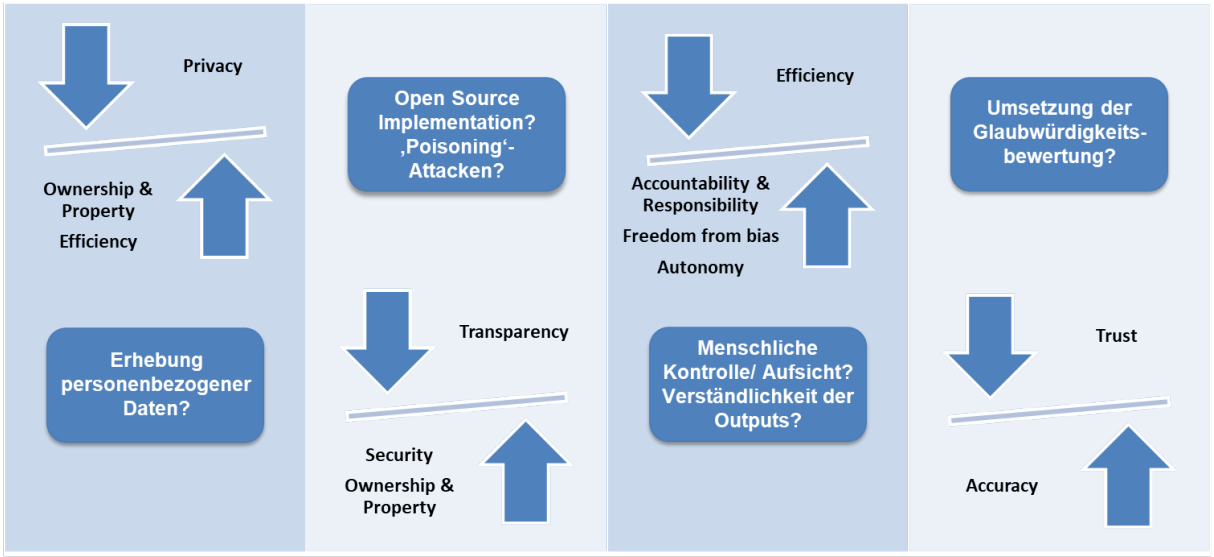

Abbildung 10. Identifizierte Werte und Wertekonflikte

Im Rahmen der Literaturstudie zeigte sich, dass OSINT Technologien sind nicht nur im Cyber-sicherheitskontext verbreitet sind, sondern eine Zunahme ähnlicher Ansätze auch in verwandten Bereichen zu verzeichnen ist. Im Bereich des Krisen- und Notfallmanagements wird OSINT unter anderem zum Zweck der Förderung eines gemeinsamen Situationsbewusstseins sowie der Verbesserung der Kollaboration, des Datenaustauschs und der kollektiven Entscheidungs-findung eingesetzt. Vor diesem Hintergrund werden auch verstärkt partizipative Design- und Technologiebewertungsmethoden diskutiert, die den besonderen organisationalen und rechtlichen Rahmenbedingungen sowie den Anforderungen an die Technologieanwendung von Einsatzkräften gerecht werden. OSINT ist dabei keine einzelne Technologie, sondern ein Rahmenkonzept, in welchem einzelne Arbeitsschritte mit unterschiedlichen technischen Ansätzen umgesetzt werden können. Im Folgenden sind die von uns erarbeiteten Implikationen für die einzelnen Teilschritte unter Berücksichtigung der Erkenntnisse aus der Literaturstudie und der identifizierten Wertkonflikte aufgeführt. Da eine Beschreibung der Wertekonflikte über den Umfang dieses Berichts hinaus gehen würde, gibt Abbildung 10 einen Gesamtüberblick über die Konflikte und die damit verbundenen und im Folgenden beschriebenen Design- und Entwicklungsherausforderungen.

- **Informationsbeschaffung (1):** OSINT-Systeme nutzen oft öffentlich zugängliche Social-Media-Daten, was Herausforderungen im Bereich Datenschutz mit sich bringt. Privacy Impact Assessments und Privacy by Design Prinzipien sollten angewendet werden, um Privatsphärenverletzungen zu vermeiden und den Datenschutz zu gewährleisten.
- **Pre-Processing (2):** Das Pre-Processing ist entscheidend, um Verzerrungen in Trainingsdatensätzen für Machine Learning (ML) zu minimieren, insbesondere bei der Erstellung von Profilen. Forschung sollte sich auf die Entwicklung geeigneter Datensätze und Leitlinien konzentrieren, um Verzerrungen im Cybersicherheitskontext zu reduzieren.

- **ML-Algorithmen (3):** Bei der Analyse von OSINT-Informationen gibt es ethische Herausforderungen bei der Auswahl von Algorithmen. Die Bedürfnisse der Nutzer hinsichtlich der Verständlichkeit und Nachvollziehbarkeit von Entscheidungen sollten berücksichtigt werden. Es besteht Forschungsbedarf zu erklärbarer KI und der Offenlegung von Quellen- und Kontextinformationen.
- **Natural Language Processing (NLP) (4):** NLP wird in OSINT-Systemen zur Erstellung kurzer, vorstrukturierter Texte genutzt. Forschung zu zielgruppenspezifischen Warnungen und Benachrichtigungen durch NLP-Ansätze ist vielversprechend, insbesondere durch den Einsatz großer vortrainierter Sprachmodelle.
- **Implementierung in Cybersicherheitskontexten (5):** Staatliche Akteure als Nutzer von OSINT-Systemen müssen erhöhte Anforderungen an Rechenschaftspflicht und Standards beachten. Es besteht ein Trade-off zwischen Effizienz und der Erfüllung dieser Anforderungen. Fallstudien könnten helfen, spezifische Richtlinien und Designanforderungen für den rechtlich zulässigen Einsatz solcher Systeme zu entwickeln.

Vor diesem Hintergrund war es ein Ziel von CYWARN, Faktoren, die mit der Akzeptanz von OSINT-Systemen für Cybersicherheitskontexte verbunden sind zu untersuchen, um eine den ethischen Anforderungen angemessene und kontext-sensible Gestaltung derartiger Systeme zu unterstützen. Im Fokus standen dabei die Einstellungen und Erwartungen der deutschen Bevölkerung als von der Technologieentwicklung- und –nutzung betroffene Stakeholder. Um die Einstellungen gegenüber der Nutzung von OSINT im Bereich der Cybersicherheit zu verstehen, haben wir eine repräsentative Umfrage in Deutschland durchgeführt. Die obenstehende Forschungsfrage wurde mit Hypothesen operationalisiert, die aus einem Review von verwandten Arbeiten abgeleitet wurden:

- H1: Personen mit einer höheren Cyber-Bedrohungswahrnehmung neigen dazu, OSINT eher zu akzeptieren.
- H2: Personen, die der Meinung sind, dass ein allgemeiner Bedarf an OSINT besteht, neigen zu einer höheren Akzeptanz von OSINT.
- H3: Personen mit größeren Bedenken hinsichtlich der Privatsphäre neigen zu einer geringeren Akzeptanz von OSINT.
- H4: Personen mit einem stärkeren Privatsphäre-Verhalten neigen zu einer geringeren Akzeptanz von OSINT.
- H5: Das Level des Bewusstseins für OSINT verändert die Zusammenhänge zwischen der Cyber-Bedrohungswahrnehmung, Privatsphäre-Bedenken, Privatsphäre-Verhalten sowie dem wahrgenommenen Bedarf an OSINT und der Akzeptanz von OSINT.

In seiner endgültigen Fassung umfasste der Fragebogen 20 geschlossene Fragen, von denen drei in der gegenwärtigen Studie ausgewertet wurden. Zur Messung der Bedrohungswahrnehmung wurden die Befragten gefragt, wie sie das Risiko einschätzen, in den nächsten fünf Jahren Opfer eines Cyberangriffs zu werden (Frage 12), und zur Messung des Datenschutzverhaltens fragten wir, wie kontinuierlich sie Sicherheitstools oder -maßnahmen auf persönlichen Geräten nutzen (Frage 13). Dann wurden die Teilnehmer gebeten, Aussagen zur Verbreitung, Nutzung und Auswirkung von OSINT (Frage 14) sowie zu OSINT-Aktivitäten von Sicherheitsbehörden (Frage 15) zu bewerten. Nach Übermittlung des Fragebogens an GapFish wurde die Umfrage programmiert und im September 2021 durchgeführt. Die Stichprobe von N=1.093 Teilnehmern wurde so ausgewählt, dass sie die deutsche Bevölkerung in Bezug auf Alter, Geschlecht, Bildung, Einkommen und Bundesland repräsentiert. Die Datenanalyse wurde mit den Softwaretools Microsoft Excel und RStudio Version 4.0.5 durchgeführt. Zunächst wurde eine deskriptive Analyse mit Antwortverteilungen für einzelne Items im Zusammenhang mit der Akzeptanz von OSINT durchgeführt. Zur Analyse der hypothetischen Zusammenhänge zwischen Cyber-Bedrohungswahrnehmung, OSINT-Bedarf, Datenschutzbedenken, Datenschutzverhalten und OSINT-Akzeptanz wurde eine multiple lineare Regression durchgeführt. Die Regression wurde zweimal durchgeführt. Während das erste Modell unabhängige Effekte darstellte, stellte das zweite, eher explorative Modell Interaktionseffekte mit dem Faktor OSINT-Wahrnehmung dar.

Die Untersuchung der OSINT-Akzeptanz mittels multipler linearer Regression ergab, dass Bedrohungswahrnehmung, OSINT-Bedarf, Privatsphäre-Bedenken und Privatsphäre-Verhalten die Akzeptanz beeinflussen. Das Modell war signifikant ($F(4,1086) = 456$, $p < .001$) und erklärte 63 % der Varianz. OSINT-Bedarf hatte den größten positiven Einfluss ($\beta = .63$, $p < .001$), während Privatsphäre-Bedenken negativ korreliert waren ($\beta = -.06$, $p < .001$). Privatsphäre-Verhalten zeigte keine signifikante Wirkung ($\beta = .02$, $p = .145$).

	Schätzung	**St. Abweichung**	**t-Wert**	**Pr(>|t|)**
Intercept	1.3783	0.0759	18.16	0.0000
Privatsphäre-Bedenken	-0.0624	0.0156	-4.00	0.0001
Privatsphäre-Verhalten	0.0205	0.0141	1.46	0.1445
Bedrohungswahrnehmung	0.0497	0.0111	4.46	0.0000
OSINT-Bedarf	0.6323	0.0156	40.46	0.0000

Tabelle 5. Regressionsmodell

Ein zweites, exploratives Regressionsmodell untersuchte die Rolle des OSINT-Bewusstseins als Interaktionsterm und fand, dass es die Zusammenhänge zwischen den Prädiktoren beeinflusst. Das Modell war ebenfalls signifikant ($F(9,1081) = 224,4$, $p < .001$) mit einem R-Quadrat von .65. Hier zeigte nur die Interaktion zwischen Privatsphäre-Bedenken und OSINT-Bewusstsein eine signifikante Vorhersagekraft ($\beta = -.06$, $p < .001$). Bei geringen Privatsphäre-Bedenken stieg die OSINT-Akzeptanz mit höherem OSINT-Bewusstsein.

	Schätzung	**St. Abweichung**	**t-Wert**	**Pr(>|t|)**
(Intercept)	0.5775	0.2265	2.55	0.0109
Privatsphäre-Bedenken	0.1406	0.0530	2.65	0.0081
Privatsphäre-Verhalten	-0.0246	0.0531	-0.46	0.6431
Bedrohungswahrnehmung	0.0061	0.0436	0.14	0.8881
OSINT-Bedarf	0.6525	0.0499	13.06	0.0000
OSINT-Bewusstsein	0.2799	0.0603	4.64	0.0000
Privatsphäre-Bedenken: OSINT-Bewusstsein	-0.0602	0.0133	-4.52	0.0000
Privatsphäre-Verhalten: OSINT-Bewusstsein	0.0103	0.0139	0.74	0.4600
Bedrohungswahrnehmung: OSINT-Bewusstsein	0.0091	0.0114	0.79	0.4275
OSINT-Bedarf: OSINT-Bewusstsein	-0.0138	0.0129	-1.07	0.2851

Tabelle 6. Regressionsmodell mit Interaktion

Die meisten OSINT-Systeme für die Cybersicherheit konzentrieren sich auf die Erkennung neuer Cyberbedrohungen und könnten zusätzliche Analysen für die Vorfallsbehandlung ermöglichen. Konzepte des kontextbezogenen Datenschutzes könnten diesen Ansatz unterstützen, z.B. als Teil begrenzter Datenerfassungsansätze. Bei der Entwicklung von OSINT-Systemen, die auf der Grundlage von CTI arbeiten, sollten daher die folgenden Implikationen berücksichtigt werden:

- OSINT-Systeme sollten die Art der gesammelten Daten berücksichtigen und Ansätze zur Datenminimierung verfolgen. In sich ändernden Bedrohungssituationen könnte die Sammlungsstrategie angepasst werden und so auf veränderte Bedrohungswahrnehmungen reagieren, die mit einer höheren OSINT-Akzeptanz verbunden sind.
- Das System sollte im sozialen Kontext verbleiben, was durch die Nutzung professioneller Quellen (Anbieter, Schwachstellen- und Cybersicherheitsexperten, ...) erreicht werden könnte, und könnte auf größere Bedrohungen reagieren, indem es über den Kontext hinaus erweitert wird, wenn sich der Bedarf ändert.
- Plattformen sollten ihre Sicherheitsvorkehrungen gegen unverhältnismäßige Datenerfassung aktualisieren und die Datensparsamkeit unterstützen.
- Das Bewusstsein der Menschen könnte durch den Einsatz partizipativer Entwurfs- und Wartungsmethoden geschärft werden.

1.5.2 Rechtlich-organisationale Implikationen (AP5.2)

Während der Projektlaufzeit wurden rechtliche und organisationale Implikationen der Projektaktivitäten analysiert. Aufgrund enger Zusammenarbeit sind die Methodik und Ergebnisse der im Abschlussbericht des HMdI hinterlegt.

2 Wichtigste Positionen des zahlenmäßigen Nachweises

Der Verlauf der Arbeit im Projekt folgte der im Projektantrag formulierten Planung. Das Vorhaben CYWARN war zunächst auf eine Laufzeit von drei Jahren angelegt. CYWARN startete formal am 01. September 2020; das Kick-off Meeting des Projekts fand am 07. Oktober 2020 statt. Trotz des erfolgreichen Meilensteintreffens am 26. April 2022 wurde deutlich, dass ein Projektpartner die noch zur Verfügung stehenden Mittel zur Projektlaufzeit nicht mehr ausschöpfen können wird. Dies ist insbesondere darauf zurückzuführen, dass während der Corona-Pandemie für die mehrfach ausgeschriebenen Stelle kein geeignetes Personal zur Umsetzung des zugehörigen Teilvorhabens des Projektpartners gefunden werden konnte. Um dennoch alle Aufgaben des CYWARN-Projekts bewältigen zu können, wurde eine Mittelreduktion des entsprechenden Projektpartners sowie eine Mittelaufstockung für andere Projektpartner durchgeführt.

Bedingt durch die Corona-Pandemie konnten die vorgesehenen Präsenzveranstaltungen nicht wie geplant durchgeführt werden; stattdessen verursachten die vermehrten Videokonferenzen erhöhte Arbeitsaufwände. Pandemie-bedingte Änderungen der Arbeitsweise führten zu vielfach erlebten Mehraufwänden und Verschiebungen. Generell zeigte sich, dass die konsortiumsweite Kooperation durch Einsatz von Telekommunikationsmitteln und Videokonferenzen gesichert werden konnte, dass jedoch die Effizienz des Austauschs und der gemeinsamen Meinungsbildung unter dem fehlenden direkten Austausch leidet. Darüber hinaus wurde festgestellt, dass die TUDa nicht alle Mittel (insb. Reisemittel) zur Projektlaufzeit fristgerecht verausgaben kann. Aufgrund der fehlenden Personalbesetzung eines Projektpartners konnten die praxisorientierte Weiterentwicklung und Evaluation der Strategien und Technologien des Projekts nicht im geplanten Maße durchgeführt werden. Daher wurde eine mittelneutrale Projektverlängerung um sechs Monate beantragt und bewilligt. Nach dem Abschlusstreffen am 18. März 2024 endete das Projekt formell am 31. März 2024.

Details sind dem begleitenden Zahlenwerk zu entnehmen.

3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die geplanten konzeptuellen Arbeiten wurden in vollem Umfang durchgeführt. Mit dem Cyber Threat Observatory wurde ein neuartiger und vielversprechender Demonstrator zur (I) automatisierten Erhebung und Integration öffentlicher und geschlossener Datenquellen, (II) intra- und interorganisationalen Datenauswertung mit Glaubwürdigkeitsanalyse der Inhalte und Quellen unter Zuhilfenahme von Visual Analytics und (III) kanalübergreifenden Kommunikation von Cyberbedrohungen an verschiedene Akteure (BürgerInnen, Behörden und KRITIS-Betreiber) insbesondere für behördliche CERTs entwickelt.

Die im Konsortium erreichten konzeptuellen und technischen Ergebnisse gehen über die Planung hinaus. So konnten außerplanmäßig KI-Algorithmen unter Nutzung von Multi-Level Finetuning, Few-Short Learning, Data Augmentation und Adversarial Examples sowie ein Sprachmodell für die Cybersicherheit umgesetzt und veröffentlicht werden. Darüber hinaus konnten drei statt der geplanten zwei Evaluationsiterationen des Cyber Threat Observatory mit CERTs durchgeführt werden. Aufgrund der fehlenden Personalbesetzung des HMdI konnten jedoch die praxisorientierte Weiterentwicklung und Evaluation der Analyse- und Kommunikationsstrategien des Projekts nicht im geplanten Maße durchgeführt werden.

Signifikantes Anwenderinteresse, positive Rückmeldungen bei Projektpräsentationen, sowie erfolgversprechende Transfer- und Verwertungsperspektiven bestätigen die Korrektheit und Relevanz der im Projekt realisierten Ergebnisse. Das Projektergebnis bestätigt damit die Sinnhaftigkeit der verfolgten Ziele und die Notwendigkeit und Angemessenheit der geleisteten Arbeit.

4 Voraussichtlicher Nutzen

Im Projekt konnten wesentliche Ergebnisse zur empirischen Fundierung, Implementierung und nutzerzentrierten Evaluation des Cyber Threat Observatory erzielt werden, welches auf die Bedürfnisse von CERTs maßgeschneidert ist.

1) Die empirische Erfassung von aktuellen Arbeitspraktiken und Anforderungen von CERTs durch Interviews (N=25) sowie der Erwartungshaltung und Wahrnehmung von BürgerInnen durch repräsentative Befragungen (N=3).
2) Entwicklung einer Basiskomponente für das Cyber Threat Observatory zur Erfassung öffentlicher Datenquellen, darunter Security Advisories, Common Vulnerabilities and Exposures, Indicators of Compromise und Social Media (AP2).
3) Methoden und Technologien auf Basis von Indikatoren und künstlicher Intelligenz zur Filterung, Bewertung und Priorisierung von Cyber-Lageinformationen, welche in die Analysekomponente des Cyber Threat Observatory eingeflossen sind (AP3).
4) Die Veröffentlichung wissenschaftlicher Ergebnisse auf Konferenzen und in Magazinen sowie die Integration von Ergebnissen in den universitären Lehrbetrieb, aber auch der Transfer für andere Anwenderzielgruppen (AP4).
5) Die Durchführung einer ethischen und sozialen Technikfolgenabschätzung sowie die Analyse rechtlicher und organisationaler Bedingungen und Implikationen (AP5).

Die erreichten Ergebnisse zielen auf die weitere Entwicklung und Erforschung im Anwendungsfeld, bieten aber auch Perspektiven für andere Szenarien, in denen die Erfassung oder Ergänzung eines Lagebilds auf Basis öffentlicher Daten relevant ist (z.B. Erkennung und Analyse glaubwürdiger und relevanter Informationen für die Reaktion auf Cybermissbrauch und in Krisensituationen). Entsprechend werden insbesondere verfolgt:

- Weiterentwicklung und Erprobung der Konzepte in Forschungsprojekten, in denen die Erfassung, Analyse und Kommunikation von Cyberbedrohungen und Sicherheitslücken relevant ist
- Kooperative Forschungs- und Entwicklungsprojekte mit industriellen Partnern zur Realisierung neuer innovativer Angebote
- Veröffentlichung weiterer Ergebnisse durch wissenschaftliche Publikationen und Open-Source-Daten (Datensätze, KI-Modelle, Programmcode)
- Beratungsdienstleistungen und Entwicklungsunterstützung für behördliche CERTs, die an einer kooperativen Weiterentwicklung von CYWARN-Ergebnissen interessiert sind und solche in ihre Lösungen integrieren wollen
- Realisierung der wissenschaftlichen Anschlussfähigkeit durch Weiterentwicklung der innovativen Ergebnisse in anschließenden F&E-Projekten mit und ohne öffentliche Förderung.

5 Fortschritte bei anderen Stellen

Das Ziel des Projekts "User-Centered Technology Design for Cyber Situational Awareness" (ATHENE-CyAware, 01.01.2023-30.12.2026, BMBF+HMWK) ist es, nutzbare Sicherheitstechnologien zu entwickeln und zu evaluieren, die das Cyber Situational Awareness von CERTs und IT-Sicherheitsbeauftragten in Unternehmen, Industrie, Regierung und Gesellschaft verbessern. Konkret umfasst dies (1) eine KI-Pipeline für die anpassbare, schnelle, leistungsfähige und transparente Erkennung von Cyberbedrohungen und -Schwachstellen sowie (2) ein visuelles Analyse-Dashboard für die Filterung relevanter Informationen. Das Projekt umfasst empirische Vorstudien zum Verständnis organisatorischer Praktiken, die Entwicklung innovativer IKT zur Bewältigung identifizierter Herausforderungen und die Bewertung ihrer Funktionalität und Gebrauchstauglichkeit. Die Ergebnisse werden in Form von Best Practices, Richtlinien und Design-Implikationen destilliert, um unser Verständnis von nutzbaren Technologien zu verbessern, die zur Cyber Situational Awareness und Entscheidungsfindung beitragen. Das Projekt wird von PEASEC umgesetzt und baut auf den Ergebnissen von CYWARN auf, indem Ansätze des Multi-Level Fine-Tuning, Data Augmentation and

Few-Shot Learning zur Verbesserung der KI-Ansätze für die Relevanzbewertung von Informationen aus öffentlichen Datenquellen, u.a. soziale Medien, herangezogen werden.

6 Erfolgte oder geplante Veröffentlichungen

1. Markus Bayer, Philipp Kuehn, Ramin Shanehsaz, Christian Reuter (2024) **CySecBERT: A Domain-Adapted Language Model for the Cybersecurity Domain** *ACM Transactions on Privacy and Security (TOPS)* ;27(2). [DOI]
2. Markus Bayer (2024) **Deep Learning in Textual Low-Data Regimes for Cybersecurity** Darmstadt, Germany: Dissertation (Dr. rer. nat.), Department of Computer Science, Technische Universität Darmstadt.
3. Marc-André Kaufhold, Thea Riebe, Markus Bayer, Christian Reuter (2024) **‚We Do Not Have the Capacity to Monitor All Media': A Design Case Study on Cyber Situational Awareness in Computer Emergency Response Teams** Proceedings of the Conference on Human Factors in Computing Systems (CHI) New York, NY, USA. [DOI]
4. Philipp Kuehn, Dilara Nadermahmoodi, Moritz Kerk, Christian Reuter (2024) **ThreatCluster: Threat Clustering for Information Overload Reduction in Computer Emergency Response Teams** arXiv. [DOI]
5. Markus Bayer, Marc-André Kaufhold, Christian Reuter (2023) **A Survey on Data Augmentation for Text Classification** *ACM Computing Surveys (CSUR)* ;55(7):1–39. [DOI]
6. Markus Bayer, Tobias Frey, Christian Reuter (2023) **Multi-Level Fine-Tuning, Data Augmentation, and Few-Shot Learning for Specialized Cyber Threat Intelligence** *Computers & Security*. doi:10.1016/j.cose.2023.103430
7. Marc-André Kaufhold, Christian Reuter, Thomas Ludwig (2023) **Big Data and Multi-platform Social Media Services in Disaster Management** In: Amita Singh: *International Handbook of Disaster Research*. Singapore: Springer Nature Singapore, 1–21. [PDF]
8. Philipp Kuehn, Mike Schmidt, Markus Bayer, Christian Reuter (2023) **ThreatCrawl: A BERT-based Focused Crawler for the Cybersecurity Domain**. [PDF]
9. Philipp Kuehn, David N. Relke, Christian Reuter (2023) **Common vulnerability scoring system prediction based on open source intelligence information sources** *Computers & Security*. [DOI]
10. Christian Reuter, Marc-André Kaufhold (2023) **Crisis Informatics** In: Zheng Yan: *Cambridge Handbook of Cyber Behavior*. Cambridge University Press. [PDF]
11. Thea Riebe (2023) **Technology Assessment of Dual-Use ICTs – How to assess Diffusion, Governance and Design** Darmstadt, Germany: Dissertation (Dr. rer. nat.), Department of Computer Science, Technische Universität Darmstadt. [DOI]
12. Thea Riebe, Julian Bäumler, Marc-André Kaufhold, Christian Reuter (2023) **Values and Value Conflicts in the Context of OSINT Technologies for Cybersecurity Incident Response: A Value Sensitive Design Perspective** *Computer Supported Cooperative Work: The Journal of Collaborative Computing (JCSCW)*. [DOI]
13. Thea Riebe, Tom Biselli, Marc-André Kaufhold, Christian Reuter (2023) **Privacy Concerns and**
25. Thea Riebe, Philipp Kuehn, Philipp Imperatori, Christian Reuter (2022) **U.S. Security Policy: The Dual-Use Regulation of Cryptography and its Effects on Surveillance** *European Journal for Security Research (EJSR)*. [DOI]
26. Stefka Schmid, Thea Riebe, Christian Reuter (2022) **Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion between Civilian and Defense R&D** *Science and Engineering Ethics* ;28(12):1–23. [DOI]
27. Markus Bayer, Marc-André Kaufhold, Christian Reuter (2021) **Information Overload in Crisis Management: Bilingual Evaluation of Embedding Models for Clustering Social Media Posts in Emergencies** Proceedings of the European Conference on Information Systems (ECIS). [PDF]
28. Rolf Egert, Nina Gerber, Jasmin Haunschild, Philipp Kuehn, Verena Zimmermann (2021) **Towards Resilient Critical Infrastructures – Motivating Users to Contribute to Smart Grid Resilience** *i-com – Journal of Interactive Media* ;20(2):161–175. [DOI]
29. Jasmin Haunschild, Christian Reuter (2021) **Perceptions of Police Technology Use and Attitudes Towards the Police – A Representative Survey of the German Population** Mensch und Computer – Workshopband. [DOI]
30. Marc-André Kaufhold, Markus Bayer, Daniel Hartung, Christian Reuter (2021) **Design and Evaluation of Deep Learning Models for Real-Time Credibility Assessment in Twitter** 30th International Conference on Artificial Neural Networks (ICANN2021) Bratislava. [DOI]
31. Marc-André Kaufhold, et al. (2021) **CYWARN: Strategy and Technology Development for Cross-Platform Cyber Situational Awareness and Actor-Specific Cyber Threat Communication** Mensch und Computer – Workshopband. [DOI]
32. Marc-André Kaufhold, Christian Reuter, Tina Comes, Milad Mirbabaie, Stefan Stieglitz (2021) **2nd Workshop on Mobile Resilience: Designing Mobile Interactive Systems for Crisis Response** *MobileHCI '21: 23nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. [DOI]
33. Marc-André Kaufhold (2021) **Information Refinement Technologies for Crisis Informatics: User Expectations and Design Principles for Social Media and Mobile Apps** Wiesbaden, Germany: Springer Vieweg. [DOI]
34. Philipp Kuehn, Markus Bayer, Marc Wendelborn, Christian Reuter (2021) **OVANA: An Approach to Analyze and Improve the Information Quality of Vulnerability Databases** Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES 2021). [DOI]
35. Tilo Mentler, Christian Reuter, Simon Nestler, Marc-André Kaufhold, Michael Herczeg, Jens Pottebaum (2021) **8. Workshop Mensch-Maschine-Interaktion in sicherheitskritischen Systemen: Ausnahmezustand** Mensch und Computer – Workshopband. [DOI]

Acceptance Factors of OSINT for Cybersecurity: A Representative Survey *Proceedings on Privacy Enhancing Technologies (PoPETs)* (1):477–493. [DOI]

14. Thea Riebe (2023) **Technology Assessment of Dual-Use ICTs – How to assess Diffusion, Governance and Design** Wiesbaden, Germany: Springer Vieweg. [PDF]
15. Elise Özalp, Katrin Hartwig, Christian Reuter (2023) **Trends in Explainable Artificial Intelligence for Non-Experts** In: Peter Klimczak, Christer Petersen: *AI – Limits and Prospects of Artificial Intelligence*. Bielefeld: Transcript Verlag, 223–243. [PDF]
16. Ali Sercan Basyurt, Jennifer Fromm, Philipp Kuehn, Marc-André Kaufhold, Milad Mirabaie (2022) **Help Wanted – Challenges in Data Collection, Analysis and Communication of Cyber Threats in Security Operation Centers** Proceedings of the International Conference on Wirtschaftsinformatik (WI) Nürnberg. [PDF]
17. Markus Bayer, Marc-André Kaufhold, Björn Buchhold, Marcel Keller, Jörg Dallmeyer, Christian Reuter (2022) **Data Augmentation in Natural Language Processing: A Novel Text Generation Approach for Long and Short Text Classifiers** *International Journal of Machine Learning and Cybernetics (IJMLC)*. [DOI]
18. Kaan Eyilmez, Ali Sercan Basyurt, Stefan Stieglitz, Christoph Fuchss, Marc-André Kaufhold, Christian Reuter, Milad Mirbabaie (2022) **A Design Science Artefact for Cyber Threat Detection and Actor Specific Communication** Australasian Conference on Information Systems (ACIS). [PDF]
19. Jasmin Haunschild, Marc-André Kaufhold, Christian Reuter (2022) **Cultural Violence and Fragmentation on Social Media: Interventions and Countermeasures by Humans and Social Bots** In: Myriam Dunn Cavelty, Andreas Wenger: *Cyber Security Politics: Socio-Technological Transformations and Political Fragmentation*. Routledge, 48–63. [PDF]
20. Jasmin Haunschild, Marc-André Kaufhold, Christian Reuter (2022) **Perceptions and Use of Warning Apps – Did Recent Crises Lead to Changes in Germany?** Mensch und Computer – Tagungsband New York. [DOI]
21. Marc-André Kaufhold, Julian Bäumler, Christian Reuter (2022) **The Implementation of Protective Measures and Communication of Cybersecurity Alerts in Germany – A Representative Survey of the Population** Mensch und Computer – Workshopband. [DOI]
22. Marc-André Kaufhold, Ali Sercan Basyurt, Kaan Eyilmez, Marc Stöttinger, Christian Reuter (2022) **Cyber Threat Observatory: Design and Evaluation of an Interactive Dashboard for Computer Emergency Response Teams** Proceedings of the European Conference on Information Systems (ECIS) Timisoara, Romania.
23. Marc-André Kaufhold, Christian Reuter, Tina Comes, Milad Mirabaie, Stefan Stieglitz (2022) **Proceedings of the 2nd Workshop on Mobile Resilience: Designing Interactive Systems for Crisis Response** Darmstadt: TUprints. [PDF]
24. Philipp Kuehn, Julian Bäumler, Marc-André Kaufhold, Marc Wendelborn, Christian Reuter (2022) **The Notion of Relevance in Cybersecurity: A Categorization of Security Tools and Deduction of Relevance Notions** Mensch und Computer – Workshopband. [DOI]
36. Christian Reuter (2021) **Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement (Zweite Auflage)** 2 ed. Wiesbaden: Springer Vieweg. [PDF]
37. Christian Reuter (2021) **Einleitung in die sicherheitskritische Mensch-Computer-Interaktion** In: Christian Reuter *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement (Zweite Auflage)*. Wiesbaden, Germany: Springer Vieweg, 3–17. [PDF]
38. Christian Reuter, Marc-André Kaufhold (2021) **Soziale Medien in Notfällen, Krisen und Katastrophen** In: Christian Reuter *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement (Zweite Auflage)*. Wiesbaden, Germany: Springer Vieweg, 407–430. [PDF]
39. Christian Reuter, et al. (2021) **Die Zukunft sicherheitskritischer Mensch-Computer-Interaktion** In: Christian Reuter *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement (2. Auflage)*. Wiesbaden, Germany: Springer Vieweg, 673–681. [PDF]
40. Christian Reuter, Marc-André Kaufhold (2021) **Informatik für Frieden-, Konflikt- und Sicherheitsforschung** In: Christian Reuter*Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement (Zweite Auflage)*. 2 ed. Wiesbaden, Germany: Springer Vieweg, , 605–630. [PDF]
41. Thea Riebe, Marc-André Kaufhold, Christian Reuter (2021) **The Impact of Organizational Structure and Technology Use on Collaborative Practices in Computer Emergency Response Teams: An Empirical Study** *Proceedings of the ACM: Human Computer Interaction (PACM): Computer-Supported Cooperative Work and Social Computing* ;5(CSCW2). [DOI]
42. Thea Riebe, Tristan Wirth, Markus Bayer, Philipp Kuehn, Marc-André Kaufhold, Volker Knauthe, Stefan Guthe, Christian Reuter (2021) **CySecAlert: An Alert Generation System for Cyber Security Events Using Open Source Intelligence Data** Information and Communications Security (ICICS). [DOI]
43. Marc-André Kaufhold (2020) **Information Refinement Technologies for Crisis Informatics: User Expectations and Design Implications for Social Media and Mobile Apps in Crisis** Darmstadt, Germany: Dissertation (Dr. rer. nat.), Department of Computer Science, Technische Unviersität Darmstadt. [PDF]
44. Philipp Kuehn, Thea Riebe, Lynn Apelt, Max Jansen, Christian Reuter (2020) **Sharing of Cyber Threat Intelligence between States** *S+F Sicherheit und Frieden / Peace and Security* ;38(1):22–28. [DOI]
45. Christian Reuter, Marc-André Kaufhold, Tina Comes, Michèle Knodt, Max Mühlhäuser (2020) **Mobile Resilience: Designing Mobile Interactive Systems for Societal and Technical Resilience** MobileHCI '20: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. [DOI]

CY-WARN
Cyber Situational Awareness & Communication

Kurzbericht des BMBF-Projekts „CYWARN“

Zuwendungsempfänger:	**Technische Universität Darmstadt (TUDa)** **Wissenschaft und Technik für Frieden und Sicherheit** **Prof. Dr. Dr. Christian Reuter** **Pankratiusstraße 2** **64289 Darmstadt**
Förderkennzeichen:	**13N15407**
Verbundtitel:	**Strategie- und Technologie-Entwicklung zur medienübergreifenden Erstellung eines Cyber-Lagebilds und akteursspezifischen Kommunikation von Cyber-Warnmeldungen (CYWARN)**
Teilvorhabentitel:	**Methoden, Strategien und Technologien zur medienübergreifenden und teilautomatisierten Cyber-Lagebilderstellung**
Laufzeit des Vorhabens:	**01.10.2020 – 31.03.2024**
Berichtsverfasser:	**Dr. Marc-André Kaufhold**
Erstellungsdatum:	**16.09.2024**

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SIFO.de

TECHNISCHE UNIVERSITÄT DARMSTADT

1 Ursprüngliche Aufgabenstellung

Das Gesamtziel von CYWARN war es, Strategien und Technologien zur Analyse und Kommunikation des *Cyber-Lagebilds* für Computer Emergency Response Teams (CERTs) zu entwickeln. Damit wurde angestrebt, die *Früherkennung* von Cybergefahren (Vorbereitung) und Umsetzung von *Gegenmaßnahmen* (Reaktion) zum Schutz von Medien- und Informationsinfrastrukturen durch CERTs zu verbessern. Der Arbeitsplan umfasst theorieorientierte systematische Literaturstudien, qualitative und quantitative empirische Studien (AP1), partizipative Demonstrator- und Strategie-Entwicklung (AP2, AP3), Evaluationen der Akzeptanz und Gebrauchstauglichkeit (AP4) und eine Abschätzung der ethischen, rechtlichen und sozialen Implikationen (AP5). Das Teilvorhaben „Methoden, Strategien und Technologien zur medienübergreifenden und teilautomatisierten Cyber-Lagebilderstellung“ des Fachgebiets „Wissenschaft und Technik für Frieden und Sicherheit“ (PEASEC) der Technischen Universität Darmstadt (TUDa) verfolgte das Ziel, unter Einbezug automatisierter Methoden CERTs in der Erhebung und Auswertung von Cyberbedrohungen und Sicherheitslücken zur verbesserten Cyber-Lagebilderstellung zu unterstützen.

2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Da technologische Entwicklungen im Rahmen der Digitalisierung, insbesondere dem Internet der Dinge und Smart Cities, für eine immer stärkere Vernetzung in privaten und professionellen Kontexten sorgen, erhöht sich ebenso die Gefahr durch Cyberbedrohungen und Sicherheitslücken in Medien- und Informationsinfrastrukturen. Zur Früherkennung von Cyberbedrohungen und Sicherheitslücken sind CERTs mit einer steigenden Anzahl öffentlicher Informationen und Quellen (z.B. Blogs, Feeds, Foto- und Videoportale, Sensoren, soziale Netzwerke und Websites) konfrontiert, die aufgrund unterschiedlicher Informationsqualität und Quellenvertraulichkeit zur Integration in ein Cyber-Lagebild gefiltert, analysiert, verifiziert und anschließend priorisiert werden müssen.

Das Projekt konnte hierbei auf grundlegende Forschung zum Situationsbewusstsein aus der Kriseninformatik zurückgreifen, die jedoch primär auf Krisen aus dem physischen und nicht dem virtuellen Raum fokussierte. Im Forschungsgebiet der Open Source Intelligence wurden Methoden zur Erfassung öffentlicher Datenquellen angeboten, die daher einen Beitrag zur Etablierung des Lagebilds im Cyberraum leisteten. Klassifikationsmethoden des maschinellen Lernens lieferten zudem eine geeignete Grundlage, um irrelevante Informationen rauszufiltern und eine Priorisierung verbleibender Cybersicherheitsinformationen zu ermöglichen. Weiterhin wurden indikatorbasierte Ansätze zur Bewertung der Glaubwürdigkeit von Informationen untersucht, die allerdings als Blackbox implementiert waren und daher von Entscheidungsträgern nicht nachvollzogen werden könnten. Der Forschungsbereich der Visual Analytics stellte dann Methoden bereit, die gesammelten, priorisierten und bewerteten Informationen filterbar und interaktiv in ein visuelles Lagebild für den Cyberraum zu integrieren. Um die ethisch-sozialen Implikationen der Forschung zu untersuchen, bestand mit dem Value-Sensitive Design eine Methode, die Werte und Wertekonflikte involvierter Stakeholder systematisch im Designprozess zu berücksichtigen.

3 Ablauf des Vorhabens

Das Vorhaben CYWARN war zunächst auf eine Laufzeit von drei Jahren angelegt. CYWARN startete formal am 01. September 2020; das Kick-off Meeting des Projekts fand am 07. Oktober 2020 statt. Trotz des erfolgreichen Meilensteintreffens am 26. April 2022 wurde deutlich, dass ein Projektpartner die noch zur Verfügung stehenden Mittel zur Projektlaufzeit nicht mehr ausschöpfen können wird. Dies ist insbesondere darauf zurückzuführen, dass während der Corona-Pandemie für die mehrfach ausgeschriebenen Stelle kein geeignetes Personal zur Umsetzung des zugehörigen Teilvorhabens des Projektpartners gefunden werden konnte. Um dennoch alle Aufgaben des CYWARN-Projekts bewältigen zu können, wurde eine Mittelreduktion des entsprechenden Projektpartners sowie eine Mittelaufstockung für andere Projektpartner durchgeführt.

Darüber hinaus wurde festgestellt, dass die TUDa aufgrund der Corona-Pandemie nicht alle Mittel (insb. Reisemittel) zur Projektlaufzeit fristgerecht verausgaben kann. Aufgrund der fehlenden Personalbesetzung eines Projektpartners konnten die praxisorientierte Weiterentwicklung und Evaluation der Strategien und Technologien des Projekts nicht im geplanten Maße durchgeführt werden. Daher wurde eine mittelneutrale Projektverlängerung um sechs Monate beantragt und bewilligt. Nach dem Abschlusstreffen am 18. März 2024 endete das Projekt formell am 31. März 2024.

4 Wesentliche Ergebnisse

Im Projekt konnten wesentliche Ergebnisse zur empirischen Fundierung, Implementierung und nutzerzentrierten Evaluation des Cyber Threat Observatory erzielt werden, einem neuartigen Demonstrator zur (I) automatisierten Erhebung und Integration öffentlicher und geschlossener Datenquellen, (II) intra- und interorganisationalen Datenauswertung mit Glaubwürdigkeitsanalyse der Inhalte und Quellen unter Zuhilfenahme von Visual Analytics und (III) kanalübergreifenden Kommunikation von Cyberbedrohungen an verschiedene Akteure (BürgerInnen, Behörden und KRITIS-Betreiber).

1) Die empirische Erfassung von aktuellen Arbeitspraktiken und Anforderungen von CERTs durch Interviews (N=25) sowie der Erwartungshaltung und Wahrnehmung von BürgerInnen durch repräsentative Befragungen (N=3).
2) Entwicklung einer Basiskomponente für das Cyber Threat Observatory zur Erfassung öffentlicher Datenquellen, darunter Security Advisories, Common Vulnerabilities and Exposures, Indicators of Compromise und Social Media (AP2).
3) Methoden und Technologien auf Basis von Indikatoren und künstlicher Intelligenz zur Filterung, Bewertung und Priorisierung von Cyber-Lageinformationen, welche in die Analysekomponente des Cyber Threat Observatory eingeflossen sind (AP3).
4) Die Veröffentlichung wissenschaftlicher Ergebnisse auf Konferenzen und in Magazinen sowie die Integration von Ergebnissen in den universitären Lehrbetrieb, aber auch der Transfer für andere Anwenderzielgruppen (AP4).
5) Die Durchführung einer ethischen und sozialen Technikfolgenabschätzung sowie die Analyse rechtlicher und organisationaler Bedingungen und Implikationen (AP5).

5 Zusammenarbeit mit anderen Forschungseinrichtungen

Das Gesamtvorhaben CYWARN wurde von PEASEC in Darmstadt koordiniert. Der gesamte Projektverbund umfasste die (1) Technische Universität Darmstadt, Wissenschaft und Technik für Frieden und Sicherheit (PEASEC), die (2) Universität Potsdam, Wirtschaftsinformatik und Digitale Transformation (DIGICAT), das (3) Hessische Ministerium des Innern, für Sicherheit und Heimatschutz (HMdI, Cyber Competence Center (Hessen3C) und die (4) Virtimo AG. Als assoziierte Partner waren darüber hinaus das (5) Ministerium des Innern und für Sport Rheinland-Pfalz, CERT-rlp, der (6) Brand- und Katastrophenschutz, Verteidigungswesen, Krisenmanagement Hessen, das (7) Landespolizeipräsidium Hessen, das (8) Bundesamt für Sicherheit in der Informationstechnik, Operative Cyber-Sicherheit, die (9) Johanniter-Unfall-Hilfe e. V. und die (10) Digitalstadt Darmstadt beteiligt.