Computer Supported Cooperative Work (CSCW)
https://doi.org/10.1007/s10606-022-09453-4

RESEARCH ARTICLE

Values and Value Conflicts in the Context of OSINT Technologies for Cybersecurity Incident Response: A Value Sensitive Design Perspective

Thea Riebe[1],*, Julian Bäumler[1], Marc-André Kaufhold[1] & Christian Reuter[1]
[1] *Science and Technology for Peace and Security (PEASEC), Technical University of Darmstadt, Pankratiusstraße 2, Darmstadt, 64289, Hesse, Germany (E-mail: riebe@peasec.tu-darmstadt.de; E-mail: julian.baeumler@stud.tu-darmstadt.de; E-mail: kaufhold@peasec.tu-darmstadt.de; E-mail: reuter@peasec.tu-darmstadt.de)*

Accepted: 22 November 2022

Abstract. The negotiation of stakeholder values as a collaborative process throughout technology development has been studied extensively within the fields of Computer Supported Cooperative Work and Human-Computer Interaction. Despite their increasing significance for cybersecurity incident response, there is a gap in research on values of importance to the design of open-source intelligence (OSINT) technologies for this purpose. In this paper, we investigate which values and value conflicts emerge due to the application and development of machine learning (ML) based OSINT technologies to assist cyber security incident response operators. For this purpose, we employ a triangulation of methods, consisting of a systematic survey of the technical literature on the development of OSINT artefacts for cybersecurity (N=73) and an empirical value sensitive design case study, comprising semi-structured interviews with stakeholders (N=9) as well as a focus group (N=7) with developers. Based on our results, we identify implications relevant to the research on and design of OSINT artefacts for cybersecurity incident response.

Key Words: Computer Emergency Response Team, Cybersecurity incident response, Machine learning, Open-source intelligence, Value sensitive design

1. Introduction

Research on Computer Supported Cooperative Work (CSCW) has driven the field of crisis informatics, which has been described as a multidisciplinary field "concerned with the ways in which information systems are entangled with socio-behavioral phenomena connected to disasters" (Soden and Palen 2018, p. 2). To respond to crises, gathering and analysing social media data for emergency ser-

Published online: 04 April 2023

vices has been studied. Especially its use for emergency operators in collaboration with informal response communities (Purohit et al., 2014), the mitigation of information overload (Kaufhold et al., 2020), and social media users' expectations towards crisis communication (Petersen et al., 2017; Reuter et al., 2017) has been explored. Similar to existing emergency services for natural disasters, Computer Emergency Response Teams (CERTs), which are also known as Computer Security Incident Response Teams (CSIRTs) serve as a central point of contact, advice, and coordination for government institutions and private actors in the event of cybersecurity incidents and threats (Kossakowski, 2001; Riebe et al., 2021a).

CERTs do not only respond to incidents, which are reported to them, they also monitor various media sources for new vulnerabilities and other threats, verify different pieces of information, analyse threats, communicate with other CERTs, and are expected to support "stakeholder[s] with specific recommendations, to provide (daily) reports for selected stakeholders (e.g., a daily vulnerability report for ministries), or to issue a general warning for multiple stakeholders (in case larger- scaled ICT infrastructures are threatened)" (Riebe et al. 2021a, p. 11). The main challenge CERTs face when executing their tasks, lies in ensuring adequate cyber situational awareness when evaluating information from numerous public and closed sources (Franke and Brynielsson, 2014; Riebe et al., 2021a). Relevant public sources such as social media, blogs, websites, and feeds can be included in this process, as part of an open-source intelligence (OSINT) approach (Glassman and Kang, 2012). Considering the risk of information overload when evaluating public sources, especially in the case of serious security incidents with many potential civilian casualties, the use of technical systems utilising machine learning (ML) algorithms for information filtering and analysis has become common (Kaufhold et al., 2020). In such decision support systems, artificial intelligence (AI) agents are becoming increasingly relevant as assistants for decision-making (Chouldechova et al., 2018). CERT members have stated that they are in need of (semi-)automated assistance for data gathering, (pre-)processing, analysis, and communication of cyber threats based on ML (Riebe et al., 2021a; Van der Kleij et al., 2017). Thus, there is the increasing use of OSINT systems within CERTs (Kassim et al., 2022). As OSINT mostly relies on private data from users of online media to be an effective tool for cybersecurity operators, the acceptance of such a system is decisive. Value conflicts may arise as a consequence of different groups in society being directly or indirectly affected in different ways, depending on the application of OSINT technologies. Therefore, it is imperative that not only OSINT systems be further researched and investigated, but also arising value conflicts. Research that focuses primarily on the values and value conflicts relevant to the development of OSINT systems for cybersecurity incident response has not yet been conducted extensively. This paper is guided by the collaborative Value Sensitive Design (VSD) method (Friedman, 1996) and will contribute to answering the following research question: **Which values and value conflicts**

emerge due to the application and development of ML-based open-source intelligence technologies in the context of cybersecurity incident response?

Our study is part of the CYWARN research project developing OSINT artefacts for CERTs (Kaufhold et al., 2021), and contributes to the CSCW-discourse with (1) a systematic literature review about technical research on OSINT technologies for the application in the domain of cybersecurity, (2) an empirically grounded elaboration of relevant stakeholder values and value conflicts in connection to the application and development of OSINT technologies for cybersecurity incident response, and (3) an outline of implications for the research on and the design of ML-based OSINT technologies for collaborative cybersecurity incident response.

This paper is structured as follows: In Section 2, related work on OSINT and VSD is presented and the research gap is outlined. Afterwards, Section 3 introduces the research design. We employ a triangulation of methods that combines an empirical case study consisting of a focus group (N=7) and semi-structured expert interviews (N=9) with a systematic literature review of technical research on OSINT technologies in the context of cybersecurity (N=73). The results of both the literature review and the empirical case study are presented in Section 4. In Section 5, insights obtained are synthesised by elaborating research and design implications and it is discussed how value sensitivity can facilitate collaboration. Finally, the limitations of the study are indicated, possible starting points for further research are outlined, and in Section 6 a brief summary of the work is provided.

2. Background and Related Work

As the design and application of novel information and communication technology (ICT) artefacts interferes with existing social practices, it is necessary to engage with the practices and problems of professionals, institutional arrangements, and technical infrastructures of the respective application environment (Wulf et al., 2011). Approaches for participatory design, which aim to address this issue, have been part of the CSCW discourse (Randall et al., 2007), as they follow the objective of facilitating cooperation (Kensing and Blomberg, 1998). Extensive research has focused on the design for collaboration in crisis response to better understand the collaborative practices and, thus, design systems which support response teams (Büscher et al., 2016; Cobb et al., 2014; Liegl et al., 2016; Reuter et al., 2014). Here, collaboration can be described as the development of a set of common practices which could be adopted by newcomers without previous participation and explanation (Heath and Luff, 1992). With regard to CERTs, this includes monitoring of and responding to cyber threats and incidents, as well as evaluating and sharing relevant information with outside parties. OSINT systems can help collaborating distributed teams to gain a shared situational awareness due to their support of context awareness, thus facilitating the establishment of a meta-perspective (Jones et al., 2021).

This section will first provide an overview on how OSINT systems as AI agents assist CERTs (Section 2.1), second introduce VSD as our participatory design approach and situate the paper in context of previous research (Section 2.2) and third, outline the research gap (Section 2.3).

2.1. OSINT Systems as AI-based Decision Support in Cybersecurity Incident Response

Central to OSINT is the idea that various pieces of publicly available information can be combined in unforeseen ways to gain innovative insights about the subject of interest (Glassman and Kang, 2012). OSINT can accordingly be defined as an activity that “involves the collection, analysis, and use of data from open sources for intelligence purposes” (Koops et al. 2013, p. 677). Approaches for cybersecurity incident response predominantly use social media as their main source (Riebe et al., 2021b), thereby taking advantage of crowdsourcing. Crowdsourcing for emergency response, however, depends on the quality and the trustworthiness of the information (Tapia and Moore, 2014).

ML algorithms are increasingly used for the automation of data gathering, pre-processing, and analysis (Williams and Blum, 2018). With the adoption of ML, challenges of explainability arise, as non-expert users are often unable to comprehend how an algorithm produces a certain output (Burrell, 2016). This is problematic as explainability is crucial to establishing users’ trust in a system (Dzindolet et al., 2003). Therefore, recent research focuses on possibilities of explainable artificial intelligence (XAI) (Longo et al., 2020; Wang et al., 2019).

As part of decision support systems, AI has gained importance in assisting teams with particular types of expertise (Bansal et al., 2019). In their study on Human-AI interaction, Zhang et al. (2022, p. 1) study “how people trust and rely on an AI assistant that performs with different levels of expertise relative to the person, ranging from completely overlapping expertise to perfectly complementary expertise”. In their experiments, they found that the “ideal partnership between humans and AI has been based on the premise of their complementary expertise” (Zhang et al. 2022, p. 20). In addition, they found that trust in AI was lowest when there was a complete overlap in the expertise of AI and human operators. Thus, trust in an AI agent is associated with the perceived usefulness of the AI and its complementary expertise. For the trust of the human operator, the style of communication of the AI agents has also been shown to be relevant (Zhang et al., 2022). As shown in an study by Feng and Boyd-Graber (2019) using human-computer teams to perform play a trivia knowledge game, the skill level of human operators is crucial for the interpretation of the expertise of AI. This is supported by Schaffer et al. (2019), who found in their study (N=529), that an AI agent was only effective at lower levels of self-assessed knowledge, whereas self-confident users often rejected the agent’s suggestions. In summary, for an effective Human-AI-Teaming in decision-making processes, the expertise of the users, the capabilities of the AI systems, e.g. managing large amounts of data in real-time and identifying similarities, as well as the communication style of the AI agents towards the users are relevant.

For cybersecurity incident response, OSINT technologies leveraging ML are primarily used in three areas. First, they are used for investigative purposes, e.g. to support digital forensics (Quick and Choo, 2018), or cyberattack attribution (Layton, 2016). Second, they are utilised for gathering cyber threat intelligence (CTI), which can be understood as "threat-related information which allows cyber security experts to investigate on a certain threat, e.g. the name of a malware, adversary or vulnerability" (Tundis et al. 2020, p. 454). Third, they are also used for risk assessment and mitigation purposes, e.g. to assess the attack surface of organisations (Hayes and Cappa, 2018), or to expose social engineering attack opportunities (Edwards et al., 2017).

In a study comprising an online survey and semi-structured interviews with staff of 13 national CERTs from Asia, Europe, the Caribbean, and North America, Kassim et al. (2022) found that the use of OSINT tools in cybersecurity incident response is on the rise. In accordance to Riebe et al. (2021a), they found that CERTs lack the resources to manage the increasing amount of public available data, which requires further verification and risk analysis. In their study on the collaborative practices of German CERTs, Riebe et al. (2021a) found that the (semi-)automation of threat detection and analysis, as well as reporting interfaces were found to be useful improvements.

2.2. VSD Research on OSINT

VSD, as a theoretically grounded method, is particularly well suited to anticipate value conflicts that arise through technology use, and proactively addresses them during design (Friedman et al., 2013). As a central theoretical assumption, VSD takes an interactional position on the relationship between technology design and social context; design features support or undermine certain values, but ultimately only their interplay with users and the context of use determines how a technology influences society (Davis and Nathan, 2015). A value can be defined as "what a person or group of people consider important in life" (Friedman et al. 2013, p. 57). VSD strives to consider direct and indirect stakeholders and their values during design (Friedman et al., 2013). As often differing values are considered important, value conflicts may arise. A value conflict exists, if competing values suggest incompatible choices as the best for the design of technical artefacts and no single value trumps all others (van de Poel and Royakkers, 2011).

In order to ensure that values are taken into account, VSD proposes a methodology that is composed of three interdependent and iteratively applied types of investigation (Friedman et al., 2013). In conceptual investigations, stakeholder groups affected by the envisaged technical artefacts are identified, and values expected to be important to them are elaborated as well as conceptualised (Friedman et al., 2013). In empirical investigations, social science methods are used to revise these findings with a focus on the opinions of stakeholders, as well as anticipated usage contexts (Manders-Huits, 2011). During both types of investigations, potential value conflicts may be identified (Friedman et al., 2013). Finally, in technical investigations, design choices that support identified and prioritised values are derived (Manders-Huits, 2011). Concerning value discovery, Le Dantec et al. (2009) argue that values should be identified during direct stakeholder engagement. In agreement with this, we utilise empirical investigations for value discovery in this work.

Several studies have specifically explored values and value conflicts in the cybersecurity domain. Among others, potential conflicts have been identified between the values security and privacy, the values security and fairness (Christen et al., 2017; Domingo-Ferrer and Blanco-Justicia, 2020; van de Poel, 2020), as well as the values security and autonomy (Christen et al., 2017; Domingo-Ferrer and Blanco-Justicia, 2020). Further, privacy was found to be potentially conflicting with both fairness and accountability (van de Poel, 2020). However, the identified conflicts mostly involve either security or privacy and altogether the works remained on a conceptual level, without reference to specific technical artefacts. Other publications referred to specific OSINT artefacts for other security purposes, but they were narrowly focused on safeguarding the value privacy through regulatory Privacy by Design approaches (Casanovas, 2017; Casanovas et al., 2014; Casanovas, 2014; Cuijpers, 2013; Koops et al., 2013; Rajamäki, 2019; Rajamäki and Simola, 2019).

2.3. Research Gap

While values and value conflicts relevant to cybersecurity have been investigated conceptually (Christen et al., 2017; Domingo-Ferrer and Blanco-Justicia, 2020; van de Poel, 2020), to the best of the authors' knowledge, there are no publications that primarily focus on the values relevant to ML-based OSINT technologies for cybersecurity incident response, despite their increasing significance. Moreover, the consideration of Privacy by Design principles (Casanovas, 2017; Casanovas et al., 2014; Casanovas, 2014; Cuijpers, 2013; Koops et al., 2013; Rajamäki, 2019; Rajamäki and Simola, 2019) has only been studied in connection to OSINT artefacts for other security related scenarios. Riebe et al. (2021b) have further shown that Privacy by Design principles are hardly taken into consideration in technical research on the development of OSINT artefacts for cybersecurity event detection. Accordingly, a research gap can be found with regard to the empirical investigation of relevant stakeholder values related to potential value conflicts resulting from the application and development of such technologies as ML-based decision support systems. The derived implications for design and research may be essential for the future development of OSINT systems for cybersecurity incident response in order to ensure their societal acceptance and stakeholder cooperation.

3. Methods

To elaborate which values and value conflicts emerge due to the application and development of ML-based OSINT technologies in the context of cybersecurity incident response, the research design uses a triangulation of methods (see Fig. 1). While the empirical investigation of relevant values and value conflicts is performed on the basis of a case study in which the results of a focus group (N=7) and of semi-structured expert interviews (N=9) are content-analysed, along with a preceding conceptual investigation of direct and indirect stakeholder groups, a systematic literature study (N=73) reviews technical research on OSINT technologies for the domain of cybersecurity. A combination of these approaches is reasonable, particularly taking into account the elaboration of the values and value conflicts can be based on an adequate empirical basis, and that it is possible to complement the gained insights with perspectives from other OSINT artefacts and application scenarios. While the methodological procedure of the literature review is described in Section 3.1, the details regarding the case study are presented in Section 3.2.

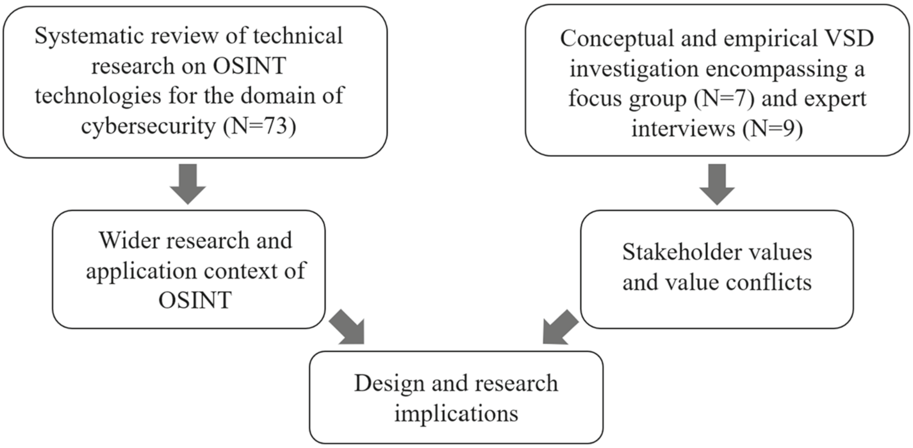
Systematic review of technical research on OSINT technologies for the domain of cybersecurity (N=73)
Conceptual and empirical VSD investigation encompassing a focus group (N=7) and expert interviews (N=9)
Wider research and application context of OSINT
Stakeholder values and value conflicts
Design and research implications

Figure 1. Illustration of the research design

3.1. Systematic Literature Review: OSINT in the Domain of Cybersecurity

To situate the findings of the empirical case study within the broader context of technical research on OSINT-technologies for application in the field of cybersecurity, the literature review section seeks to answer the **following questions**:

1. For which deployment scenarios in the cybersecurity domain are OSINT technologies being developed?
2. What technical features, techniques, and data sources are used?
3. Are ethical, legal, and social implications taken into consideration?

As this review follows an explicit and reproducible method to identify and evaluate the publications, it can be considered a systematic literature review (vom Brocke et al., 2015). Specifically, a sequential review approach is used in which literature search, analysis, and the writing of the review follow a step-by-step process (Levy and Ellis, 2006). As research conducted by private actors and state bodies in many cases is not accessible, only research published in academic publications is taken into account. For the review, a search in the literature databases ACM Digital Library, IEEE-Xplore, Science Direct, and Springer Link was conducted. As the review focuses on technical research, the selection of the databases was based on their coverage of computer science literature and the number of publications they contain. Moreover, to ensure the quality of the reviewed works, it seemed sensible to limit the search to publications in peer-reviewed journals and conference proceedings. Finally, only work published from the beginning of the databases' coverage to the end of May 2021, the beginning of the literature research, was included. The full-text and metadata search in the databases was conducted with the following search expression using Boolean operators: *("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").* The procedure of publication search and selection is illustrated in Table 1.

The search resulted in 1,419 preliminary results, of which 945 were papers published in journals and conference proceedings. In a next step, the articles' abstracts were screened to identify irrelevant publications to the goal of the review. First,

Table 1. Procedure of the publication selection for the systematic literature review, differentiated by database

Database	Initial results	Journals & Proceedings	Relevant publications
IEEE-Xplore	409	356	44
ACM Digital Library	155	147	8
Springer Link	569	313	16
Science Direct	286	136	5
Total	1,419	945	73

publications not focused on the development of OSINT artefacts for the cybersecurity domain, including those related to cybercrime in a broader sense, were excluded. Second, publications in which the processing of publicly available data is only a secondary aspect of an artefact were excluded. Third, research published in languages other than English was excluded. Finally, inaccessible papers and duplications were excluded. This resulted in the exclusion of 872 publications. The remaining 73 were quantitatively analysed with Excel. A table with the categories and subcategories of analysis can be found in Appendix A. The categories were compiled in response to the three guiding questions of the review. A structured examination of the usage scenarios, features, technical approaches, and data sources of available OSINT approaches, as well as their attention to ethical, legal, and social implications (ELSI), is crucial to derive design and research implications that extend beyond the individual case studied in depth in this paper. The subcategories were initially generated by screening review papers and chapters on OSINT (Pastor-Galindo et al., 2020; Simran et al., 2020; Tundis et al., 2020). They were then revised in light of a preliminary engagement with the selected publications before the final analysis was performed.

3.2. Conceptual and Empirical VSD Case Study

3.2.1. Conceptual Stakeholder Analysis

A conceptual investigation helped to identify the stakeholders directly and indirectly affected (Friedman et al., 2017). For this purpose, a structured workshop was conducted within the research project team, in which, building on potential use cases, it was asked which groups interact with or are affected by OSINT artefacts. The results are presented in Section 4.2. In a next step, the authors identified potential harms and benefits for stakeholders as well as potentially implicated values and established working definitions based on relevant literature (Friedman et al., 2013).

3.2.2. Data Collection: Focus Group and Semi-structured Interviews

In order to identify relevant values and value conflicts, we conducted a focus group within the team of developers and researchers and nine semi-structured interviews with key stakeholder groups. In designing the procedure for data collection, we adapted the approach of Mueller and Heger (2018). Table 2 summarises the interviews and the focus group conducted.

The focus group (F1) involved seven participants from the fields of computer science, media and cognitive sciences, and software development, who were all part of the CYWARN research project, including one staff member of a German state level CERT. The sample consisted of six male participants and one female participant. The design of the focus group followed the recommendations by Krueger and Casey (2015). The discussion was held digitally and was semi-structured by a moderation guideline. After an input about VSD and a hypothetical usage scenario of the OSINT artefacts in development as a stimulus to facilitate a discussion, we asked the participants to brainstorm and write down ethical, legal, and social implications on a digital board. Afterwards, we went through the issues collected and asked the participants to discuss them with a focus on potentially implicated stakeholder values and value conflicts.

The semi-structured expert interviews (Gläser and Laudel, 2010; Kallio et al., 2016) were designed to gather empirical insights on the values important to key stakeholder groups. To collect the data, we followed a convenience sampling approach and sent interview requests to relevant organisations and individuals. When selecting the participants, we drew on the insights of the stakeholder analysis (see Subsection 4.2) and took care to involve both stakeholders directly and indirectly affected by technology development; however, since indirect effects may be experienced by a wide array of actors, we restricted the scope to stakeholders that might be most significantly affected (Friedman et al., 2017). Overall, we interviewed nine individuals from three stakeholder groups: (1) Five interviews were conducted with CERT employees, as they belong to the prospective user group of the developed artefacts. (2) Three interviews were conducted with further potential users as it is intended to transfer the artefacts to other application domains as well (I6, I7, I8). Specifically, we interviewed information security officers of a state company (I6) and a humanitarian organisation active in disaster relief (I7), as well as the head of a virtual operations support team (VOST) (I8). (3) Finally, to consider the perspective of individuals potentially affected by OSINT gathering, we interviewed one individual who is regularly disseminating cybersecurity information on social media and is active in cybersecurity related civil society organisations (I9). After obtaining the interviewees' informed consent, several blocks of questions were asked based on an interview guideline, which was slightly adapted to suit the particularities of the different stakeholder

Table 2. Overview of the interviews and the focus group with the involved stakeholder groups and the respective types of organisations

No.	Type	Stakeholder	Organisation
I1	Interview	Direct Users	State CERT
I2	Interview	Direct Users	State CERT
I3	Interview	Direct Users	State CERT
I4	Interview	Direct Users	State CERT
I5	Interview	Direct Users	University CERT
I6	Interview	Potential Users	State company
I7	Interview	Potential Users	Humanitarian organisation
I8	Interview	Potential Users	Civil protection VOST
I9	Interview	Affected by Data Collection	Civil society
F1	Focus Group	Developers & Researchers	Public university, software development company, state CERT

groups. The interview sessions were conducted online, were recorded and lasted on average 74 minutes.

3.2.3. Data Analysis: Qualitative Content Analysis

After the focus group and the interviews were transcribed, a software-assisted and category-based structuring qualitative content analysis following (Kuckartz, 2016) was conducted. We worked with thematic categories that were developed deductively on the basis of existing literature on values, as well as inductively during the analysis of the empirical material. In this study, the main category *Value* with ten subcategories, as well as the main category *Value Conflict* were used. The categories were defined in a codebook and supplemented with coding rules and examples. A shortened version of the codebook can be found in Appendix B. The transcripts were coded with the qualitative content analysis software MAXQDA. First, all the material was revised to select coding examples for each category. Then, the focus group and two interviews were coded to verify the intercoder agreement with MAXQDA. This resulted in a kappa coefficient after Brennan and Prediger (1981) of 0.69, what can be interpreted as a good result (Rädiker and Kuckartz, 2019). Furthermore, the codebook was later revised in order to further increase intercoder agreement. The text segments assigned to each category were then assembled and analysed together.

4. Results

In the following, the results of the literature study are presented in Section 4.1. Afterwards, Section 4.2 introduces the stakeholder groups identified and outlines the results of the content analysis of the empirical material.

4.1. OSINT-Technologies in the Domain of Cybersecurity

Of the 73 publications evaluated in the review, 10% named investigative purposes as the intended **scenario of use** for the systems. In 74% of the publications, systems were developed for primary use in the context of gathering CTI, in 12% for use in the area of risk assessment and mitigation, and in 4% for both investigative and CTI purposes. The temporal distribution of publications per year is shown in Fig. 2.

The publications were also examined concerning respective **features** of the systems. In 44 publications, data gathering methods were either an integrated part of the artefacts, or new data sets were specifically created in the context of research. In 36 publications, approaches for the detection of cybersecurity events have been developed. This included models for the detection of emerging cybersecurity topics (Al-Ramahi et al., 2020; Dalton et al., 2017; Kawaguchi et al., 2017; Schäfer et al., 2019), the aggregation of individual pieces of information into security events (Alves et al., 2019; Alves et al., 2021; Azevedo et al., 2019;

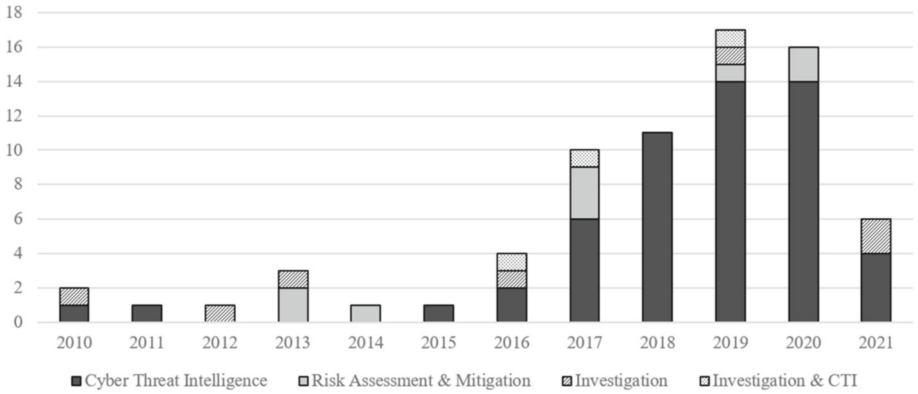

Figure 2. Number of publications per year, differentiated by intended scenarios of use

Vacas et al., 2018), the detection of distinct types of information (Behzadan et al., 2018; González-Granadillo et al., 2019; González-Granadillo et al., 2021; Liao et al., 2016; Syed, 2020), and the detection of threats related to specific infrastructures (Dionisio et al., 2019) or products (Kannavara et al., 2019; Neil et al., 2018; Nunes et al., 2018). Approaches to the classification or filtering of relevant information are presented in 26 publications, and 15 systems comprise data visualisation functions, including Social Network Analysis to explore relationships in hacker forums and marketplaces (Huang et al., 2019; Huang and Ban, 2019; Schäfer et al., 2019).

While twelve systems have the capacity to generate reports or structured pieces of information, e.g. Indicators of Compromise (IoCs), eleven systems aim to identify specific users or communities. This is related to the assessment of organisational attack surfaces or penetration testing (Chitkara et al., 2020; Edwards et al., 2017; Urban et al., 2020), the identification of individuals with insider threat potential (Kandias et al., 2013a; Kandias et al., 2013b; Kandias et al., 2013c; Kandias et al., 2017), and the investigation of hacker forums and marketplaces (Fallmann et al., 2010; Huang et al., 2019; Huang and Ban, 2019; Schäfer et al., 2019). Finally, five papers demonstrate techniques to analyse the quality or credibility of CTI (Ghazi et al., 2018; Gong et al., 2018; Jo et al., 2021; Khurana et al., 2019; Liu et al., 2017), and three propose methods to assess the quality or credibility of CTI sources (Gong et al., 2018; Liu et al., 2017; Tundis et al., 2020).

Additionally, the publications were analysed for the use of selected **algorithmic approaches** (see Fig. 3).

Most frequently, in 45 cases, algorithms for classification were implemented. Clustering, on the other hand, was only used ten times and regression only once. In addition, 13 papers used named-entity recognition, i.e. the classification of named entities in unstructured text into predefined categories for the purpose of information extraction, and seven papers used latent Dirichlet allocation for topic modelling, i.e. the discovery of previously undefined topics in a document corpus.

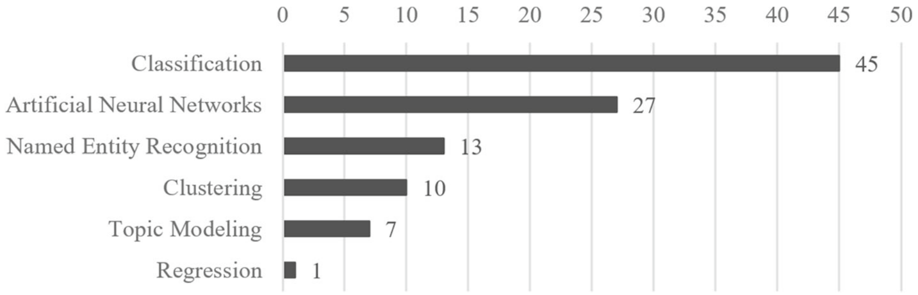

Figure 3. Algorithmic approaches implemented in the artefacts developed

Artificial neural networks were used in 27 systems. Concerning the use of ML, 46 systems used supervised ML, 28 unsupervised ML, one semi-supervised ML and 19 none. In line with the features of the examined OSINT systems, the research is focused on ML algorithms that assist operators in managing the high volume, variety, and velocity of big data by using trained classifiers, self-learning neural networks, named entity recognition, clustering, topic modeling, and regression to identify cybersecurity events, threats, and threat actors, as well as to assess the relevance, quality, or credibility of CTI and respective sources.

A variety of different **sources of information** were used with the systems.

Twitter was used 20 times, followed by cybersecurity blogs, forums, or websites that were utilised eleven times. Information from hacker forums, as well as from CTI feeds and platforms was accessed ten times each. Information from other social networks, e.g. Reddit, Facebook, and YouTube, was processed in nine instances, while seven systems made use of data gathered from dark web forums and marketplaces. Less common data sources can be found in Fig. 4.

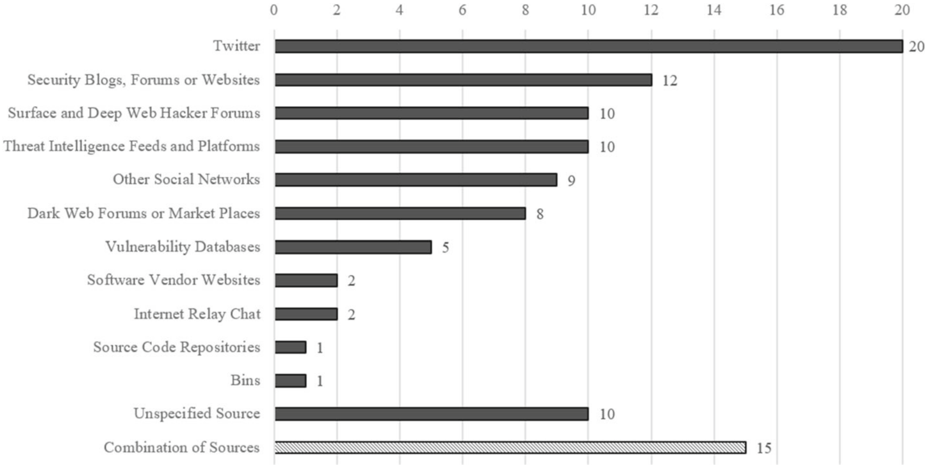

Figure 4. Publicly available information sources used for data gathering with the artefacts developed

Finally, it was examined whether **ELSI** of the respective systems were discussed. Of the 73 papers, only eleven considered such issues. While some authors argued that using only publicly available data circumvents ethical issues (Pournouri et al., 2019; Pournouri and Akhgar, 2015), Edwards et al. (2017) justified their decision not to list individuals in reports on organisations' social engineering attack surface with the concern that this could cause disciplinary action. In addition, to increase algorithmic comprehensibility, they decided to use decision tree classifiers to identify employee profiles. In a similar study, Urban et al. (2020) emphasised strict compliance with data protection requirements and the avoidance of any legally or ethically questionable strategies for data acquisition. With regard to the investigation of dark web marketplaces, Lawrence et al. (2017) mitigate the risk of legal ramifications by restricting web scraping to cybercrime related sections, textual data, and non-personal information. Ranade et al. (2018) motivated their development of a deep learning model for CTI translation partly on the premise that analysts are often not allowed to use third party services due to privacy, security, and confidentiality policies. Beyond that, a trade-off between data protection and demands of forensic investigators to have access to proactively collected data is discussed by Nisioti et al. (2021). The most extensive discussion of ELSI is found in the context of research on the identification of employees with insider threat potential. Negative effects on personal and human rights of those affected, as well as dangers concerning algorithmic profiling are discussed, and the recommendation that such screenings should be subject to strict preconditions is provided (Kandias et al., 2013a; Kandias et al., 2013b; Kandias et al., 2013c; Kandias et al., 2017).

4.2. Stakeholder Values and Value Conflicts

During the preliminary conceptual investigation, six main stakeholder groups affected by the application and development of OSINT artefacts in the domain of cybersecurity incident response were identified. Figure 5 presents the stakeholder groups and their interaction with OSINT artefacts.

The first stakeholder group consists of the individuals that interact directly with OSINT systems. In the context of this case study, these are the employees of CERTs who are expected to use a demonstrator with OSINT components. The second stakeholder group comprises actors that are indirectly affected by the collection of publicly available data with OSINT systems. In this case study, these include, in particular, actors that disseminate information on threats on social media. While the third stakeholder group, the direct beneficiaries, is directly advised and supported by the direct users of OSINT artefacts - in the case of CERTs primarily public authorities, critical infrastructure operators, and enterprises - the fourth group, the indirect beneficiaries, only receives unidirectional communication about threats and best practices - in the case of CERTs, among others, citizens and other cybersecurity organisations. The fifth stakeholder group,

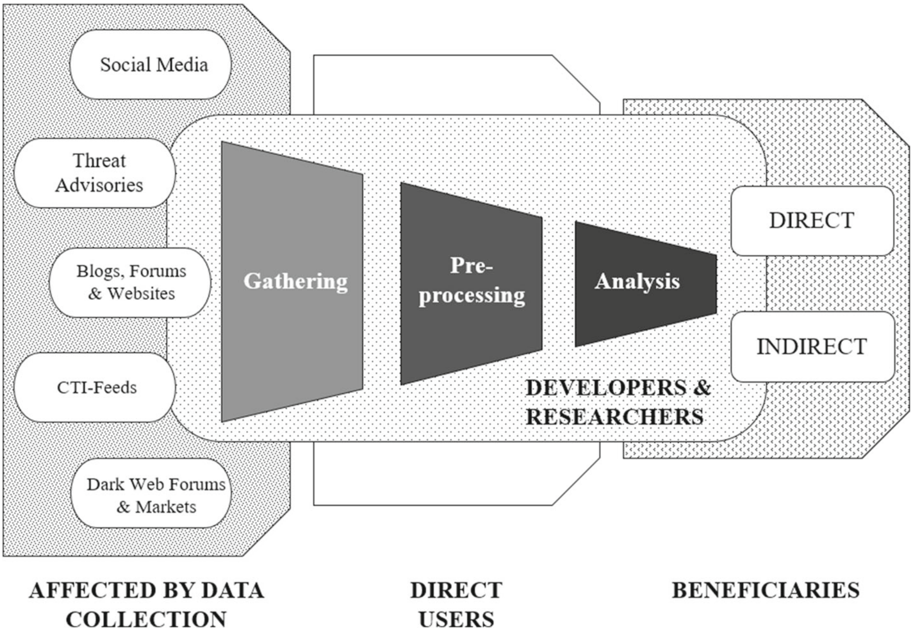

Figure 5. Main stakeholder groups of the prospective OSINT framework and their interaction with OSINT artefacts

the potential users, comprises actors that have an interest in using OSINT systems. In our case study, there may be both potential users in the field of cybersecurity and in other domains, e.g. law enforcement, civil protection, and emergency services. Finally, the developers and researchers concerned with OSINT artefacts comprise the sixth stakeholder group. In our case, this encompasses individuals from both academic research and private software engineering.

4.2.1. Stakeholder Values

During the content analysis, ten values were identified. Table 3 shows which values were discussed in the individual interviews and the focus group, and how often they were coded in total.

Accuracy can be defined as the correspondence or closeness of a statement or piece of information to the truth, the reality, or a differently defined standard (Hayes et al., 2020). Accuracy is particularly relevant in connection to ML algorithms and the quality of data. CERT staff, potential users, and developers emphasised the importance of the accuracy and quality of different types of data. The accuracy of data collected was considered very important (I1, I3, I4, I5, I7, I8, F1). Gathered information should not only be correct, but also structured consistently and have minimal redundancy to enable effective analysis (I1, I4). Since this requires repetitive and time-consuming activities, interviewees suggested drawing on the expertise of ML algorithms to harmonise information from

Table 3. Overview of the identified values and the number of coded sections

Value	I1	I2	I3	I4	I5	I6	I7	I8	I9	F1	$\sum$
Accuracy	X	X	X	X	X	X	X	X	X	X	69
Security	X	X	X	X	X	X	X	X	X	X	68
Efficiency	X	X	X	X	X	X	X	X	X	X	67
Accountability & Responsibility	–	X	X	X	X	X	X	X	X	X	27
Autonomy	X	–	–	X	X	–	X	X	–	X	27
Transparency	X	–	–	X	X	–	–	X	X	X	27
Privacy	X	–	–	X	X	X	X	X	X	X	25
Ownership & Property	–	X	–	X	X	X	–	–	X	X	22
Freedom from Bias	X	–	X	X	–	–	X	–	X	X	21
Trust	X	–	–	–	X	X	X	X	X	X	19

X signifies that a value is present in an interview or the focus group

heterogeneously structured texts and aggregate multiple pieces of information related to the same topic (I1, I4). Furthermore, the issue of disinformation was highlighted: “You also have to be cautious not to be fooled by people who try to make themselves important and publish something that is not true” (I1). The output data of algorithms needs to be relevant for OSINT analysis (I1, I4, I6, I8, F1), as well as for the information requirements of clients (I1, I2, I4, I5, I6, I7, I8, F1). For these reasons, specifically the accuracy of algorithmic decisions and the quality of training data for ML algorithms were highlighted (I5). Yet, it was argued that the application of ML should be limited to very specific tasks, as human expertise is crucial for creative or unstructured activities:

> “ML-supported systems ... are built for pattern recognition and the patterns are trained. And you just have to get out of the pattern thinking, which is really thinking inside a box” (I5).

The interviewee potentially affected by data collection pleaded for a reduction of biases in algorithms (I9). This was also emphasised by the developers with a view on algorithms for prioritisation and credibility assessment of CTI (F1).

At a high level of abstraction, **security** can be conceptualised as “the state of being free from danger or threat” (van de Poel 2020, p. 50). CERT employees and potential users highlighted the importance of security in relation to the IT infrastructure of organisations in their area of responsibility and the data processed by clients (I1, I2, I3, I4, I5, I6, I7). To ensure security, OSINT is used to leverage the expertise of numerous cybersecurity experts (I1, I3, I5, I6, I7, I9). Their expertise lies in detailed and up-to-date knowledge of specific cyber threats (I1, I9), threat actors and their strategies (I1, I9), vulnerabilities (I1, I6, I7, I9), and protection and mitigation measures (I6, I7). The civil society representative called for OSINT tools to be operated in a secure environment (I9). Finally, the developers also addressed the security of the ML algorithms against poisoning attacks, especially if information about training data and algorithmic models used is publicly accessible:

> “If a hacker notices something like this, that in some form [data] is merged and recommendations are derived from it, ... he can carry out a targeted attack based on it” (F1).

Efficiency describes the ability to accomplish specific tasks or outputs with minimal expenditure of resources (Cousins et al., 2019). In the interviews with CERT staff and potential users, efficiency considerations were cited as a key rationale for the intention to use OSINT tools (I1, I2, I3, I4, I5, I6, I7, I8). Furthermore, the efficiency gain may also improve the quality of certain services:

> “If the data collection process is simplified, then it will be intensified on the other side. Because if I am relieved of the data collection, then the evaluation will probably be more intensive. Then I might take a much closer look at the reports, which I might have published before with the watering can principle” (I6).

Specifically, possible efficiency gains were identified through technical support in the acquisition and evaluation of security advisories (I6, I7, I8), the evaluation of cybersecurity websites and blogs (I1, I3, I6, I7), the search of Twitter and other social networks for cybersecurity-relevant information (I6, I7), and supporting communication by providing target-specific cybersecurity reports or alerts (I1, I2, I3, I4, I6). Particularly for the extraction of information from unstructured texts, the use of ML algorithms has been suggested (I1, I8, I9). Here, the expertise of ML-based information extraction techniques, is to discover specific pieces of information in unstructured texts or to create summaries (I1, I9). The developers saw an interest in efficiency gains through OSINT tools also among the direct beneficiaries of CERT activities, who could receive faster support in case of incidents (F1). Finally, with a view on development, it was also suggested to keep in mind that it should be as easy as possible to adapt the artefacts to changing legal requirements (F1).

Accountability can be seen as “the (moral) obligation to account for what you did or what happened (and your role in it happening)” (van de Poel 2011, p. 39). In contrast, **responsibility** is directed towards current actions and their prospective consequences, as it refers to the obligation to evaluate one’s own role and duties in relation to a situation or a context of action (van de Poel and Royakkers, 2011). CERT staff members pointed out that alerts and reports must be approved by superiors for reasons of political accountability. (I2, I3, I4). In particular, a fixed approval process for alerts hinders automation: “There are too many sensitivities or responsibilities involved to automate something like this” (I2). With regard to disaster management, the importance of documenting verification steps and analysts involved was also pointed out in order to render the evaluation of information comprehensible for decision-makers (I8). Referring to CERTs' use of OSINT tools, the interviewee from civil society pleads for a responsible protection of the data infrastructures used (I9). It was also pointed out that when processing certain data, the design of OSINT tools should consider the obligations for CERTs to comply with reporting chains and guidelines (F1). In this context, the question was raised to what extent clear responsibilities for the consequences of incorrect predictions of ML algorithms can be ensured:

> "So if security vulnerabilities are perhaps not taken seriously, even though they are announced on social media, because this relevance algorithm has perhaps decided that it is irrelevant for various reasons, there would also be the question of whether CERTs would perhaps even be legally liable in some way, because they should actually have acted" (F1).

With regard to ICT, **autonomy** can be understood as users' ability to control the technical systems in a context appropriate manner, and to enable decisions deemed suitable for them to achieve their objectives (Friedman and Kahn, 2002). The consideration of the autonomy of stakeholders was brought forward by direct users, potential users, and developers. One interviewee in particular places the value at the centre of human-computer interaction:

> "So really the point is that you don't have to replace anyone in that sense, but you can support everyone. So I see the point with all technology that it should still be supportive, it should be a tool for people. But it should not determine people" (I5).

The complete automation of analytical OSINT processes with the help of ML is seen particularly critical, as "artificial intelligence logic always trims someone down to blinkered thinking and an increasingly narrow focus" (I5), thus restricting the analysts' evaluative capabilities. Furthermore, ensuring the autonomy of users was also discussed in the context of the adaptability of the selection of sources (I1) and the relevance assessment of information (I4, I5). For the latter, an evaluation by experienced analysts was considered crucial (I4, I5). Potential users also advocated for a prioritisation of information that could be individually adapted to the respective infrastructure (I7, I8).

Transparency can be best understood in relation to a situation in which it is beneficial for actors to make knowledge and information about a certain topic extensively available, accessible and comprehensible, without obscuring any information (Turilli and Floridi, 2009). A CERT employee advocated for the disclosure of contextual information on algorithmic decisions of OSINT artefacts to analysts (I5). Similarly, a potential user reported that the degree of transparency of algorithmic decisions should always depend on the expertise and task of the respective user group, as too much information can also be counterproductive, especially in time-sensitive situations (I8). The developers discussed the promises and pitfalls of open sourcing the code of the OSINT artefacts to be developed

(F1), while our interviewee from civil society requested transparency on the part of the developers and, ideally, an involvement of the cybersecurity community in the development of OSINT artefacts:

> “So of course I would be happy if the whole system is open source as far as possible, subject to this evaluation and the risks, and is also open development. So it’s not just open source, here’s the software. But open development” (I9).

For this work, **privacy** can be defined as “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others” (Westin 1967, p. 7). The importance of privacy was raised by CERT employees in conjunction with compliance with the legal requirements of data protection legislation (I1, I4, I5). In particular, the automated analysis of personal data is legally problematic and sometimes only granted with special permission (I1). Thus, “in the ideal case, the data ... is completely without personal reference” (I1). In the interviews with potential users it became clear that organisations are subject to very different regulations regarding privacy and data protection (I7, I8). The respondent from the group of those potentially affected by data collection considered the protection of private data a central principle: “Well, I would generally have a stomach-ache with it, if it was private data. So not publicly available data” (I9). The developers discussed privacy aspects of the development of the OSINT artefacts with a focus on the principles of data minimisation, the necessity of a justification for storing data, requirements on data deletion and anonymisation, as well as the adaptability of artefacts to changing legal requirements (F1).

According to Friedman and Kahn (2002), the value **ownership and property** is related to the rights of individuals or groups to possess, use, manage, derive profit from, or bequeath objects or pieces of information. For CERT employees and the developers, questions of ownership and property are important when it comes to legal requirements regarding the extent of data collection and the type of data to be collected (I1, I4). One CERT employee describes that the e-government law of the respective state strongly affects the processing of personal data, which should also be taken into account in the design of OSINT artefacts (I1). One potential user expressed the view that organisational policies on data processing may need to be changed before OSINT tools can be applied (I6). In addition, a part of the focus group discussion focused on the question of who should have the right to use the artefacts:

> “Perhaps it would be conceivable for a government to somehow offer the tool ... to make it available as open source and that even the public can somehow co-develop it or use it” (F1).

The value **freedom from bias** is associated with the absence of systematic unfairness against individuals or groups (Friedman and Kahn, 2002). Both the CERT staff and other organisations’ employees stressed the importance of addressee-oriented communication that is free from any systematic bias (I3, I4, I7). Pre-formulated templates for alerts were mentioned as a possible solution to this issue, because “if you have different stakeholders with technical skill levels, you can relatively easily find the right tone” (I4). Furthermore, when distributing warnings for a broad target group, appropriate communication channels should be chosen (I7). Specifically with regard to the use of ML in OSINT systems, our interview partner from civil society warned against the tendency to systematically replicate a pre-existing bias in training data:

> “The problem such systems always have is that, whatever framing or bias exists in the data and structures, machine learning ... will simply consider it as a relevant parameter” (I9).

During the focus group it was raised that the algorithmic credibility assessment of information sources may have detrimental consequences, if the labelling of an actor as an untrustworthy source became public or lead to permanent non-inclusion in future analyses (F1).

For the purpose of this paper, **trust** may be understood as “expectations, assumptions or beliefs about the likelihood that another’s future actions will be beneficial, favorable or at least not detrimental to ones’ interests” (Robinson 1996, p. 576). For direct stakeholders, trust in respective providers of information plays a major role in the verification of information from public sources (I1, I5). The developers, however, discussed trust in context of the societal acceptance of the use of OSINT technologies (F1). The trust of citizens in those using such systems may be influenced by the transparency towards the public:

> “But perhaps trust in general also depends very much on who operates the tool in the end, whether the whole thing is transparent, i.e. how much is communicated about the artificial intelligence to the outside world, what data is collected (F1).

4.2.2. Value Conflicts

While engaging with the stakeholders, eight value conflicts arised. These are illustrated in Fig. 6 together with the associated design issues.

Privacy conflicts first emerge between the privacy of actors affected by data collection and the value of ownership and property in terms of the requirements for CERT staff to be allowed to use non-anonymised data with reference to individuals (F1, I1, I4). While respect for the privacy of data subjects requires refraining from collecting personal data, it may be of interest for CERTs to collect such information. “So we’re pretty restricted there, and I think if you develop us a tool that we use in the CERT, it’s subject to those same regulations” (I1), stated a CERT employee. Thus, besides the ethical weighing of both values, the consideration of privacy and data protection requirements is central, e.g. when determining what data is collected or whether personal data is minimised, anonymised, or

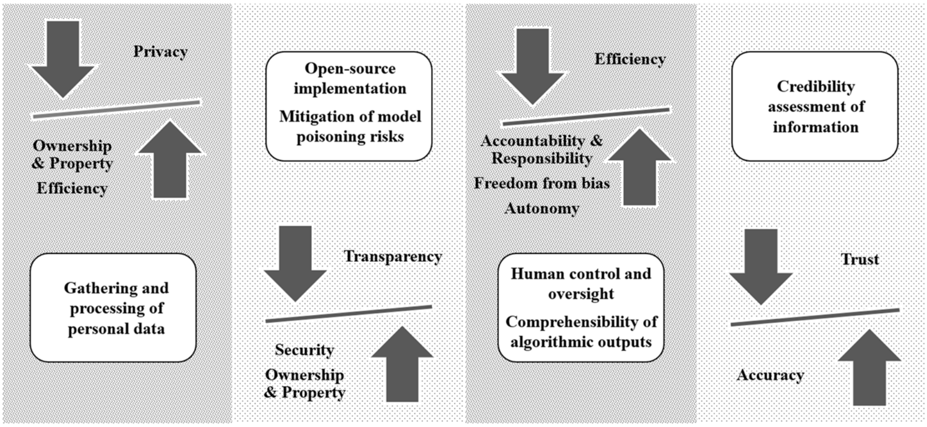
Privacy
Ownership & Property
Efficiency
Gathering and processing of personal data
Open-source implementation
Mitigation of model poisoning risks
Transparency
Security
Ownership & Property
Efficiency
Accountability & Responsibility
Freedom from bias
Autonomy
Human control and oversight
Comprehensibility of algorithmic outputs
Credibility assessment of information
Trust
Accuracy

Figure 6. Value conflicts and associated design issues identified in the empirical material

deleted (F1, I1, I4). Demands of safeguarding privacy and compliance with data protection regulations also partially conflict with the value of efficiency on the part of the CERT staff (I1). Semi-automated aggregation and analysis of public information is a key requirement of CERTs that would come with time savings, yet it was pointed out that data protection requirements might prohibit such functionalities: “This automated evaluation of public sources is not permitted to all CERTs, some of them are not allowed to do this for legal reasons” (I1).

Transparency conflicts arise, as the interviewee potentially affected by data collection, in particular, demands transparency about the specifications of the technical artefacts, training data, and ML algorithms used in, as well as the scope of data collected with OSINT tools (I9). Developers suggested that such transparency-motivated decisions could be counterproductive to the value of security, in terms of ensuring the reliable functioning of the ML models:

> “If it is known from which sources learning has taken place, one has of course again... you obviously provide an attacker the opportunity to poison the models. To do this model poisoning” (F1).

In connection to a prospective open-source implementation, a possible conflict with the value of ownership and property on the part of the developers of OSINT tools was brought forward: “You might not want to disclose the training data or explain the algorithms in detail so that you can still earn money commercially with it” (F1).

The interviews and the discussion revealed three different **efficiency conflicts**. First, due to the use of ML to accelerate OSINT processes, a conflict with the values of accountability and responsibility might arise. Considering the stakeholders whose data is processed, and the actors who receive information from CERTs, it is imperative for information to be correct, guidelines to be adhered to during processing, and misconduct to be clearly attributed to responsible actors (I2, I3, I4). ML algorithm based decision-making could undermine accountability, but conversely, the integration of manual control steps could imply higher resource consumption (I2, I3, I4, F1). Moreover, the question to what extent liability for algorithmic errors may be allocated to CERT personnel is unresolved:

> "If vulnerabilities are not taken seriously, despite being announced on social media, because this relevance algorithm has perhaps decided that it is irrelevant, there is the question whether CERTs might somehow be liable" (F1).

Second, due to the utilisation of ML algorithms, a conflict could arise between efficiency and freedom from bias. This especially applies to the direct and indirect beneficiaries of generated alerts. Warning messages generated by algorithms should be adapted to the target group to avoid systematic discrimination (I4). This, however, "means that it takes a lot of effort to reach the right level of communication" (I4), thus coinciding with a higher consumption of resources during development and application. Since CERT members expressed concerns that the present state-of-the-art is not sufficient to automatically generate target group-specific alerts (I2, I4), it seems appropriate to split the communication step into two individual tasks, thereby leveraging the expertise of both ML-based natural language processing (NLP) techniques as well as CERT analysts. In a first step, efficiency in communication could be enhanced by using NLP models to generate text segments based on a set of threat scenario- and target group-related parameters (I2, I4). In a second step, the expertise of analysts is employed to adapt the text to ensure that it actually reflects the status, requirements, and expertise of the target audience (I2, I4, I5, I9). Thereby, it is ensured that bias in communication is limited.

Third, a conflict between efficiency and users' autonomy emerges. It is particularly important for users of OSINT artefacts to remain in control over technical processes (I1, F1). However, it was highlighted that "many of the points that are aimed at, for example, additional manual control would significantly increase the time it takes for decisions to be made and solutions to be developed" (F1), thus resulting in a lower efficiency. Conversely, an exclusive focus on resource-saving optimisation may diminish operators' autonomy. Trade-offs arise especially at the stages of the design process when it is determined which decisions should be handed over to ML algorithms and to what extent users should be able to supervise these decisions. An adequate balance between both values is particularly important for OSINT tasks, where the expertise of ML algorithms and CERT analysts complement each other and can thus yield advantages over exclusively manual or automated solutions. In our context, this is especially the case with the relevance and credibility assessment of CTI. While the strength of ML in relevance assessment lies in a rapid evaluation of large amounts of information using predefined relevance criteria (I1, I6, I7), analysts can draw on this to select actually relevant information using their contextual knowledge about serviced infrastructures, e.g. deployed software (I1, I4, I5, I6, I7). During credibility assessment, three types of expertise may interact. While the expertise of ML algorithms is to compute a credibility rating using features of information previously evaluated as credible or non-credible (I5, I8), analysts, taking into account the rating and underlying contextual information, supplementary research and personal experience, as well as, if necessary, the opinions of external experts, can ultimately verify a piece of information (I1, I3, I5, I7, I8, I9). Whereas for these two tasks the trade-off between autonomy and efficiency can be mitigated by a two-step procedure, interviewees advocated for a non-automated criticality evaluation of vulnerabilities, hence prioritising autonomy (I1, I4, I6). Here, analysts resort to the expertise of external experts, which lies in their ability to determine the general criticality on basis of detailed knowledge about affected hardware, software, or corresponding exploits (I1, I4, I5). This evaluation, which can be reflected in a rating according to the Common Vulnerability Scoring System, enables analysts to decide, on basis of knowledge of the serviced infrastructure, whether there is a necessity to prioritise the vulnerability (I1, I4, I5, I6).

An **accuracy conflict** involving the value trust became apparent in the context of the credibility assessment of CTI. Both interviewed CERT staff and potential users emphasised the importance of trust in the respective providers for the selection and verification of information (I1, I5, I7, I8). In this context, trustworthiness is determined based on respective sources' past reliability (I1, I5). However, it was pointed out "that only the trustworthy position of a communication partner does not of course ensure that he does not publish nonsense anyway" (I1). Thus, in the development of ML algorithms for credibility assessment, an exclusive consideration of characteristics of trustworthy sources could compromise the accuracy of the output data.

5. Discussion and Implications

To answer the research question: **Which values and value conflicts emerge due to the application and development of ML-based open-source intelligence technologies in the context of cybersecurity incident response?** this paper has investigated the state of technical research on OSINT technologies for cybersecurity, as well as stakeholders, values, and value conflicts relevant for their application in the field of cybersecurity incident response. In this section, implications for the design of OSINT systems for this domain and for research are elaborated (Section 5.1). This is followed by a discussion on how sensitivity to the uncovered values and value conflicts can facilitate collaboration (Section 5.2), as well as an outline of the study's limitations and opportunities for future work (Section 5.3).

5.1. Research and Design Implications

The use of OSINT increases in many domains (Pastor-Galindo et al., 2020). In the area of emergency management, OSINT is used for the purpose of crisis response and shared situational awareness and collaboration (Akhgar et al., 2013; Backfried et al., 2012; Bernard et al., 2018), data sharing (Skopik et al., 2016; Mtsweni et al., 2016), and collective sense-making (Büscher et al., 2018). This has led to an increased discussion of participatory design and technology assessment methods which account for the specific organisational and legal characteristics and technology use of emergency management organisations (Büscher et al., 2018; Liegl et al., 2016). OSINT is not a single technology, but a framework in which individual steps can be performed with various technical approaches. In all three steps envisioned in Fig. 7, ML algorithms can be used. While they can support the extraction, deduplication, and harmonisation of cybersecurity information during data gathering and pre-processing, they can also contribute to the relevance and credibility assessment of CTI in the following analysis phase. Finally, in terms of communication, they can be used to pre-formulate warning messages as a foundation for their customization by CERT staff to fit the respective target groups. In the described steps, human and ML-expertise can be complementary and in interaction increase the effectiveness of CERTs. However, this study also identified values and value conflicts that need to be considered when designing OSINT technologies for cybersecurity incident response. In the following, implications for the individual OSINT steps will be discussed, while taking the findings of the literature survey and the identified value conflicts into account.

The systematic literature review revealed that **information gathering (1)** is mostly conducted using publicly available data from social media platforms. As personal information is shared on such platforms, surveyed stakeholders have

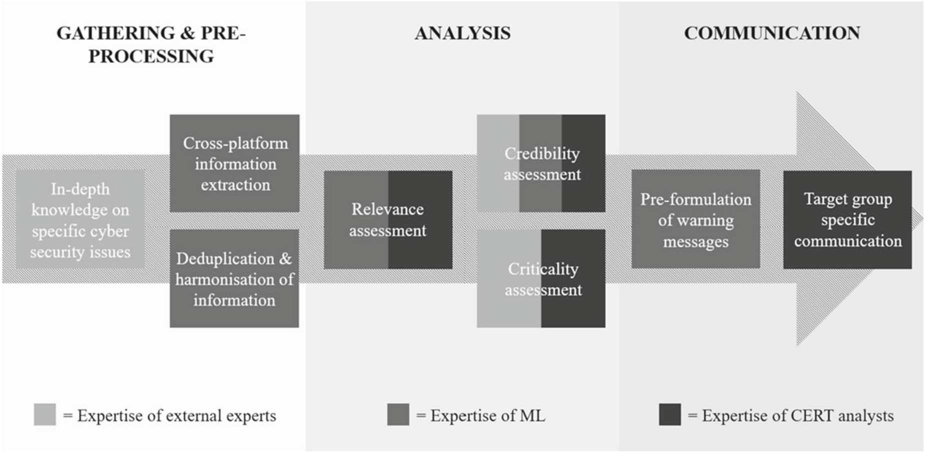
GATHERING & PRE-PROCESSING
ANALYSIS
COMMUNICATION
In-depth knowledge on specific cyber security issues
Cross-platform information extraction
Deduplication & harmonisation of information
Relevance assessment
Credibility assessment
Criticality assessment
Pre-formulation of warning messages
Target group specific communication
= Expertise of external experts
= Expertise of ML
= Expertise of CERT analysts

Figure 7. Human and ML expertise in the cybersecurity OSINT process

indicated that challenges arise in connection to privacy protection and compliance with data protection regulations. Therefore, Privacy Impact Assessments specifically for OSINT in the context of cybersecurity are needed (Liegl et al., 2016; Wright and Friedewald, 2013). The extent to which privacy infringement can be prevented by exclusively using sources specialised on the distribution of cybersecurity related information should be further analysed (Riebe et al., 2021b). In an evaluation of the Cyber Threat Observatory dashboard with CERT-employees, Kaufhold et al. (2022) found that the modular and customisable integration of different data sources and feeds has been identified as a crucial feature. With regard to information extraction from long unstructured texts, ML approaches offer clear advantages over the performance of human analysts. Specifically, their expertise lies in topic discovery and information summarisation. Since interviewees emphasised that the use of such ML techniques would increase efficiency and consequently enable the gathering of a larger amount of data for subsequent steps, their use can be recommended. For a summary of the observations and the derived implications see Table 4.

The **preprocessing (2)** of gathered information is a sensitive part of the system, as biases in the data used to train algorithms might be detected in this stage (see Table 5). Serious consequences may occur if an artefact's objective is to infer human characteristics and relationships or to profile individuals, as was the case with some of the artefacts described in the publications of the literature review. However, none of these publications discussed issues of bias and potential countermeasures. Stakeholders' demands to minimise bias in training datasets for ML algorithms as part of OSINT systems should therefore be addressed in research through studies on the creation and evaluation of appropriate datasets, the development of guidelines for the inclusive annotation of training data, and the establishment of guidelines for the evaluation and documentation of training datasets (Friedman and Hendry, 2019). With regard to cyber threat data, the guidelines would need to reflect the cybersecurity context, the respective data source, and potential bias related human as well as other characteristics. Another

Table 4. Observations and design implications for information gathering

Key Observation	Design and Research Implications
(1) Sources not specialised on cybersecurity such as social media are utilised more frequently than cybersecurity specific sources	(1) Examine legal requirements relevant to data gathering
(2) Privacy vs. Efficiency: Interest in collecting data collides with requirements of proportionality and event-relatedness	(2.1) Utilise cybersecurity specific sources (2.2) Implement data minimisation and data deletion intervals
(3) ML outperforms human analysts in extracting information from long unstructured texts	(3) Use ML-based information extraction techniques for topic discovery and information summarisation

Table 5. Observations and design implications for preprocessing

Key Observations	Design and Research Implications
(1) No discussion of bias in training data in the reviewed publications	(1) Develop guidelines to understand and limit bias in datasets
(2) Structuring gathered data in a coherent way and reducing redundancies is time-consuming	(2) Use ML techniques for information harmonisation (e.g. named entity recognition) and redundancy reduction (e.g. clustering)

challenge, according to our interviewees, is to structure collected data in a consistent way and reduce redundancies prior to further analysis. Since associated tasks are repetitive and time-consuming, they suggested drawing on the expertise of ML algorithms, which lies in harmonising information from heterogeneously structured texts (e.g. named entity recognition) and in grouping multiple pieces of information on the same topic together (e.g. clustering), thus reducing the amount of redundant information.

Implications for the development of ML algorithms arise in connection to the **analysis (3)** of OSINT information (see Table 6). While the literature review showed that algorithms are used for a variety of tasks, algorithm selection was rarely reflected from an ethical or social point of view, with exception of a publication that justifies the selection of a decision tree classifier by improving the comprehensibility of algorithmic decisions (Edwards et al., 2017). The empirical investigation showed that value conflicts can occur when algorithm selection disregards operators' needs regarding the comprehensibility, traceability, and influenceability of algorithmic decisions. With respect to the selection and development of algorithms that meet end-users' requirements, there is a need for further research on exploring the applicability of XAI and white-boxing

Table 6. Observations and design implications for analysis

Key Observation	Design and Research Implications
(1) Efficiency vs. Autonomy: Safeguarding human control and oversight may restrict scope and efficiency of data analysis	(1.1) Examine applicability of algorithmic white-boxing solutions to models for cybersecurity purposes (1.2) Give operators possibility to adapt algorithmic decision-making
(2) CERT analyst and ML expertise overlap during relevance and credibility assessment	(2) Implement two-step procedure that enables analysts to definitively assess relevance and credibility based on algorithmic pre-assessments
(3) Accuracy vs. Trust: Relying exclusively on characteristics of trustworthy sources may impair the accuracy of algorithmic credibility assessment	(3.1) Include features of pieces of information in credibility assessment (3.2) Disclose criteria used for credibility assessment to system operators

approaches for OSINT and the evaluation of different algorithmic solutions with end-users, e.g. by considering the recommendations for XAI by (Wang et al., 2019), which include support reasoning and hypothesis-generation, as well as access to source and situational data. During the interviews, it became apparent that ML can support analysts primarily in relevance and credibility assessment. As shown by Zhang et al. (2022), ML algorithms with complementary expertise are most useful to human operators. However, since in ML-assisted relevance and credibility assessment the human and algorithmic expertise overlap on a specific task, it is particularly important to ensure that human and algorithmic steps are clearly delineated by design so that advantages and limitations of both can surface. This can be achieved by implementing a two-step procedure in which the analyst always makes the final decision on the basis of an algorithmic pre-assessment, under disclosure of relevant decision parameters. In addition, as indicated by research on human-AI interaction experiments, understanding the parameters of algorithmic decisions will be crucial to establish system operators' trust (Feng and Boyd-Graber, 2019; Schaffer et al., 2019).

In the literature review, we found that NLP techniques are used in many systems, but with regard to the **generation of alerts and text (4)**, this is limited to the creation of pre-structured texts such as IoCs (see Table 7). It seems worth investigating whether NLP approaches can also be used for the generation of target group specific alerts and notifications. Advances in fundamental NLP research, especially in conjunction with the development of large pre-trained language models, might be leveraged for the development and training of models for these specific cases. However, the use of such models must be seen in light of the tension between the values efficiency and freedom from bias. In order to streamline communication while ensuring that warnings and notifications do not disadvantage relevant target groups, it is advisable to implement a two-step process. In a first step, large pre-trained language models can swiftly generate text segments based on a few parameters. In a second step, analysts can draw on their knowledge and experience of the needs and proficiency of target groups to adapt the texts accordingly. This mitigates the tension and leverages the complementary expertise of NLP models and CERT analysts, potentially increasing confidence in the system (Zhang et al., 2022).

With regard to OSINT systems' **implementation into the context of cybersecurity incident response (5)** (see Table 8), some of the reviewed studies

Table 7. Observations and design implications for communication

Key Observation	Design and Research Implications
(1) Applications of NLP for text generation are limited to the creation of pre-structured texts	(1) Harness advancements in NLP research for the generation of target specific cybersecurity alerts
(2) Efficiency vs. Freedom from bias: Algorithmic generation of warnings and notifications may reduce target group specificity	(2) Manually adapt NLP generated text segments to ensure target group specificity of warnings and notifications

raised questions of accountability and responsibility in connection with consequences of processing illegal material (Lawrence et al., 2017), or compliance with organisational secrecy and security regulations (Ranade et al., 2018). However, the challenge that state actors are often subject to enhanced requirements in terms of safeguarding accountability and compliance with different standards and responsibilities, which were also emphasised by consulted stakeholders, remained unaddressed. Since considering such requirements results in a higher resource consumption and may prevent the utilisation of particular ML algorithms, a trade-off with the value efficiency occurs. Thus, the challenge lies in developing OSINT systems that support the documentation of the operators' decisions without disproportionately impairing efficiency and usability. It is advisable to conduct case studies on the specific requirements of respective governmental user groups with regard to ensuring accountability, clear responsibilities, and reporting chains, and based on this derive concrete guidelines for the legitimate application of OSINT systems as well as requirements for their design. Finally, in the empirical investigation, stakeholders voiced a demand for transparency on training data used for ML algorithms and on OSINT system specifications, which, in turn, may increase opportunities for model poisoning and, thus, conflict with safeguarding the security of ML models. While first studies have proposed solutions to mitigate this threat (Khurana et al., 2019; Longo et al., 2020), there is a need for continued research on the magnitude of the problem and technical countermeasures. With regard to the reconciliation of transparency and security, the involvement of stakeholders in a scrutiny committee that reviews algorithm design could be a reasonable solution.

Table 8. Observations and design implications for the implementation of the OSINT system into the CERT context

Key Observation	Design and Research Implications
(1) Transparency vs. Security: Demands for transparency on system capabilities may increase security risks (e.g. model poisoning)	(1) Examine model poisoning risks and possible mitigation measures
(2) Efficiency vs. Accountability & Responsibility: Prerequisites of governmental organisations to link the processing and analysis of OSINT data to human decision-making in order to ensure accountability could impair system speed and efficiency	(2.1) Involve stakeholders in scrutiny committee that reviews algorithms (2.2) Conduct case study research on specific accountability requirements and reporting chains of governmental user groups (2.3) Develop guidelines for a legally compliant use of OSINT systems

5.2. Value Sensitivity as a Facilitator of Collaboration

Understanding value conflicts is not an end in itself, but offers venture points for value-sensitive technology design and detailed evaluations of conflicts in complex socio-political systems. From a CSCW perspective, three implications for supporting multi-actor collaboration emerge from the findings of this research paper: First, as the work of CERTs strongly relies on collaboration with other CERTs, authorities, and organisations (Riebe et al., 2021a), a tool for shared situational awareness needs to be trustworthy and support the operators reasoning and sense-making (Ley et al., 2014; Lukosch et al., 2015). Trust can be achieved by supporting the operators alignment with legal provisions and social norms. As OSINT systems work with different ML algorithms, research on the explainability of the systems and on solutions to maintain the autonomy of the operators are crucial in all application domains. Second, with regard to the communication of cyber threats, CERTs need to collaborate with different stakeholders to improve their situational awareness and provide risk mitigation strategies. It became apparent that bias-free and addressee-specific communication is pivotal to fulfilling these objectives, a factor also to be taken into account in the design of systems with communication functionalities. Additionally, the spread of social media, in particular, has opened up opportunities for CERTs to leverage novel resources. However, this paper also highlights the challenges and concerns of how this information is used and processed in such a demanding and time-sensitive collaborative environment. Therefore, the results of this study can be of use for the field of control room research, e.g. in the context of traffic management (Jones et al., 2021) or other emergency services (Normark and Randall, 2005). Third, OSINT, especially when using social media as sources, is dependent on information provided by the respective medium's users. Therefore, it involves the use of crowdsourcing, which is collaborative (Liu, 2014). Social media users need to trust OSINT operators using their data (Tapia and Moore, 2014), which can be achieved by ensuring transparency and accountability, e.g. by organisational oversight infrastructures, as well as data minimisation by Privacy by Design approaches.

5.3. Limitations and Future Work

The findings of this work must be considered in the light of some limitations, which at the same time, however, offer impulses for future research. First, the empirical investigations in this study were limited to selected stakeholder groups. In addition, only one individual potentially affected by data collection was interviewed. Thus, to consolidate the findings, further qualitative interviews and focus groups are necessary. For enquiries about citizens' attitudes, however, quantitative surveys appear to be more suitable. Our future research will therefore also include a representative survey on the attitudes of the German population towards the use of OSINT technologies. Second, the generalisability of the results is limited due to the case study design of the VSD-approach. However, similar cases of

ML-based OSINT systems for cybersecurity can utilize the design implications. Within this limitation, this work pursued the goal of elaborating values and value conflicts as abstractly as possible. Nevertheless, as the interviews and the discussion were strongly focused on the design of OSINT systems for aggregating CTI for the CERT context, the results are primarily relevant with regard to artefacts for this application field. Accordingly, studies focusing on systems for investigation and risk assessment and mitigation purposes represent promising avenues for further research. Third, this work only includes conceptual and empirical VSD investigations. In the further course of our project, it is therefore intended to conduct technical VSD investigations to derive concrete design requirements and find technical solutions through which value conflicts are minimised and preferred stakeholder values are supported as adequately as possible.

6. Conclusion

In this paper, we employed a triangulation of methods to investigate which values and value conflicts are relevant to the application and development of ML-based OSINT technologies in the context of cybersecurity incident response. In order to situate our empirical findings in the broader research and application context, we first systematically reviewed the technical research literature on the development of OSINT artefacts for the cybersecurity domain (N=73). Then, an empirical VSD case study, comprising semi-structured interviews (N=9) and a focus group (N=7) for data collection, including a subsequent qualitative content analysis of the gathered material, was undertaken to identify values of key stakeholders and to systematise potential value conflicts. The results of the literature review underlined the identified research gap, as despite research activities on OSINT for cybersecurity have increased, stakeholder values and other ethical, legal, and social issues have only been addressed in a minority of publications. In the empirical investigation, we identified ten values and eight value conflicts, particularly involving privacy, transparency, efficiency, and accuracy, that are relevant to the application and development of OSINT artefacts for cybersecurity incident response. Drawing on our findings, we derived implications for the design of and research on ML-based OSINT technologies for this application domain and discussed how sensitivity to the uncovered value conflicts and the division of tasks between human operators and ML algorithms can facilitate collaboration. Though certain limitations remain, this paper offers a systematic review of the technical research literature on the development of OSINT technologies for cybersecurity (C1), an empirically grounded elaboration of values and value conflicts related to the development and application of OSINT technologies for cybersecurity incident response (C2), and an elaboration of research and design implications for ML-based OSINT technologies for collaborative cybersecurity incident response (C3).

Appendix A: Category System of the Literature Review

The category system was created for the structured quantitative analysis of the selected publications in the literature review (see Table 9).

Table 9. Categories and subcategories used for the quantitative analysis in the literature review

Category	No.	Subcategory
Year	1	–
Country	2	–
Usage Scenario	3-a	Investigation
	3-b	Cyber Threat Intelligence
	3-c	Risk Assessment and Mitigation
	3-d	Several
Technology Features	4-a	Data acquisition
	4-b	Relevance Classification or Filtering
	4-c	Quality/ Credibility/ Reliability assessment of CTI
	4-d	Quality/ Credibility/ Reliability assessment of CTI Sources
	4-e	Cyber Event Detection
	4-f	Data Visualisation
	4-g	Report/IoC Generation
	4-h	Social Network Analysis
	4-i	Assessment of Organisational Attack Surface
	4-j	Identification of Communities or Users
Algorithmic Approaches	5-a	Classification
	5-b	Regression
	5-c	Clustering
	5-d	Topic Modelling
	5-e	Named Entity Recognition
	5-f	Artificial Neural Network
Machine Learning Type	6-a	Supervised learning
	6-b	Unsupervised Learning
	6-c	Semi-supervised Learning
	6-d	None
Data Sources	7-a	Vendor Websites
	7-b	Security Blogs, Forums, or Websites
	7-c	Dark Web Forums or Market Places
	7-d	Surface and Deep Web Hacker Forums

Table 9. (continued)

Category	No.	Subcategory
	7-e	Threat Intelligence Feeds and Platforms
	7-f	Vulnerability Databases
	7-g	Twitter
	7-h	Other Social Networks
	7-i	Source Code Repositories
	7-j	Internet Relay Chat
	7-k	Bins
	7-l	Unspecified Source
	7-m	Combination of Sources
Consideration of	8-a	Consideration
ELSI Aspects	8-b	No Consideration

Appendix B: Codebook

In the qualitative content analysis following Kuckartz (2016), sections that are covered by the definitions and coding rules of several categories can be coded with multiple categories. Coding units are sections of meaning describing one coherent thought. A shortened version of the codebook is presented below (see Table 10).

Table 10. Codebook used for the qualitative content analysis

Category	Definition	Example
Value	In the section, one or several values, i.e. abstract concepts deemed important to people's lives (Friedman et al., 2013), are described as either important or desirable to individuals, groups, and organisations; or are referred to as relevant for the design of technical artifacts, systems, or processes.	"That it could also depend a bit on transparency. For example, whether you want to make the training models comprehensible to the public in the sense of open source or something. So how their social media data is evaluated, so to speak, and for what purpose, and... all that kind of thing. So transparency maybe as a value here" (F1).
Value Conflict	In the section, either a conflict between two or more competing values that suggest two or more different and incompatible choices for the design of technical artefacts, systems or processes is described (van de Poel and Royakkers, 2011).	"And there is of course the tension between, for instance, whether an author's name is displayed or not. It can, so to speak, contain information in order to conduct further research, but perhaps it must be anonymised for data protection reasons, which of course is also an area of tension between knowledge interest and data protection" (F1).

Table 10. (continued)

Subcategory	Definition	Example
Accuracy	In the section, accuracy, i.e. the correspondence or closeness of a statement or piece of information to the truth, the reality or a differently defined standard (Hayes et al., 2020), is highlighted as an important, relevant, or desirable property of the training, input or output data of technical artefacts, systems or processes, as well as of the information communicated by CERTs.	"Well, of course we only try to subscribe to information that is relevant to us or to our target groups, as we always call them. So we have a pretty good overview of the software that is in use in the state, of course not in detail. So we would never subscribe to Security Advisories for an obscure product that probably nobody has in use" (I1).
Security	In the section, security, i.e. the state of a referent object that is free from threat or danger (van de Poel, 2020), is stressed as important, relevant, or desirable.	"First of all, the whole thing would have to be set up in a secure environment that is operated according to the ISMS and the state of the art. If I install something here and run it in an unsecured cloud environment and anyone can break into it or get access to it, that might not be a good idea. That means it should be an environment that is really provided with resources for IT security and active operation" (I9).
Efficiency	In the section, efficiency, i.e. the ability to accomplish tasks or achieve outputs with the minimal expenditure of resources (Cousins et al., 2019), is stressed as an important, relevant, or desirable property of technical artifacts, systems, and processes, or characteristic of individuals, groups, or organisations.	"Of course it would be good if information from outside could already be adequately processed, automated and sorted. The more it is automated, the more it takes work off your hands, logically. That would be all well and good, of course, but I couldn't tell you now in what form that would be feasible. But technical support, for me as a computer scientist, I am always available in any form" (I2).

Table 10. (continued)

Accountability & Responsibility	In the section, accountability, i.e. the obligation to face the consequences of and account for previous actions and decisions related to a technical artifact, system or process (van de Poel, 2011), or responsibility, i.e. the obligation to evaluate one's own role and duties with regard to a situation or a context of action (van de Poel and Royakkers, 2011), is stressed as important, relevant, or desirable.	"What are the consequences if the AI makes wrong predictions? That is the question. Okay then maybe in the end there will be no report. So I just basically asked myself what can be the consequences of all this. Can this also go so far that it somehow has legal consequences, that someone is somehow held responsible for the fact that in the end the result was not achieved that should have been achieved" (F1).
Autonomy	In the section, preserving or enhancing the autonomy of actors, i.e. their ability to independently formulate and pursue their objectives, exert their evaluative capability and implement their decisions with minimal external restrictions in a given external context (May, 1994), is stressed as important, relevant, or desirable.	"So I see with all technology the point that should still be supportive, it should be a tool for the human being. But it should not determine the people. And I think that is a very big point" (I5).
Transparency	In the section, transparency, i.e. a situation or state that is beneficial to the knowledge of individuals, groups, or organisations about a topic or fact related to a technical artefact, system or process (Turilli and Floridi, 2009), is stressed as important, relevant, or desirable.	"So of course I would be happy if the whole system is open source as far as possible, subject to this evaluation and the risks, and is also open development. So it's not just open source, here's the software. But open development" (I9).
Privacy	In the section, respecting privacy, i.e. individuals', groups', or organisations' claim to determine for themselves when, how, and to what extent information about them is communicated to others (Westin, 1967), and the demand that the flow of personal information adheres to legal and social norms (Nissenbaum, 2009), is stressed as important, relevant, or desirable.	"Are CERTs legally allowed to collect and store social media data? So that's a legal question, actually. And perhaps also which data then specifically from social media, so really everything that is there or somehow, yes certain things then also not that are somehow private or so" (F1).

Table 10. (continued)

Ownership & Property	In the section, ownership and property, i.e. the rights of individuals, groups or organisations to possess, use, manage, profit from or bequeath objects or pieces of information (Friedman and Kahn, 2002), are stressed as important, relevant, or desirable.	"On the one hand, from a legal point of view, the first issue is data collection. Because you always need a basis and an argument for what the data is collected and who is allowed to collect the data. Of course, the question is on what basis it happens and how generally this basis can be expanded. Because when I collect the data for the first time, I don't necessarily know what's in it" (I4).
Freedom from Bias	In the section, freedom from bias, i.e. the absence of any systematic discrimination against individuals or groups (Friedman and Kahn, 2002), is stressed as important, relevant, or desirable.	"What we've been thinking about is actually whether we should have some kind of news feed or something, where people can subscribe to it proactively. But even there we have the situation, of course, if people don't subscribe to something like that, then you don't reach the people. So I think our fundamental problem is how do I adequately reach people? Which medium is best?" (I7).
Trust	In the section, trust, i.e. expectations, assumptions or beliefs of actors about the likelihood that other actors' future actions will be beneficial, favourable or at least not detrimental to them (Robinson, 1996), are stressed as important, relevant, or desirable.	"But perhaps trust in general also depends very much on who operates the tool in the end, whether the whole thing is transparent, i.e. how much is communicated about the AI to the outside world, what data is collected" (F1).

Appendix C: Abbreviations

Table 11 provides an overview of the abbreviations used and the corresponding complete terms.

Table 11. List of abbreviations used in the paper

Abbreviation	Complete Term
AI	Artifical intelligence
CERT	Computer Emergency Response Team
CSCW	Computer Supported Cooperative Work
CTI	Cyber threat intelligence
ELSI	Ethical, legal, and social implications
ICT	Information and communication technology
IoC	Indicator of Compromise
HCI	Human-Computer Interaction
ML	Machine learning
NLP	Natural language processing
OSINT	Open-source intelligence
VOST	Virtual operations support team
VSD	Value Sensitive Design
XAI	Explainable artificial intelligence

Acknowledgements

This research was supported by the German Federal Ministry for Education and Research (BMBF) in the project CYWARN (13N15407) and by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Compliance with Ethical Standards

Conflict of Interests

The authors declared that they have no conflict of interest.

VSD Study of OSINT for Cybersecurity Incident Response

Open Access

References

Akhgar, Babak; Dave Fortune; Richard E. Hayes; Bárbara Guerra; and Marco Manso (2013). Social media in crisis events: open networks and collaboration supporting disaster response and recovery. *HST: 2013 IEEE international conference on technologies for Homeland security, Waltham, MA, USA, 2013*. New York: IEEE, pp. 760–765.

Al-Ramahi, Mohammad; Izzat Alsmadi; and Joshua Davenport (2020). Exploring hackers assets: topics of interest as indicators of compromise. *HotSoS'20: Proceedings of the 7th symposium on hot topics in the science of security, Lawrence, Kansas, USA, 2020*. New York: ACM, pp. 1–4.

Alves, Fernando; Aurélien Bettini; Pedro M. Ferreira; and Alysson Bessani (2021). Processing tweets for cybersecurity threat awareness. *Information Systems*, vol. 95, pp. 1–18.

Alves, Fernando; Pedro Miguel Ferreira; and Alysson Bessani (2019). Design of a classification model for a twitter-based streaming threat monitor. *DSN-w: 49th annual IEEE/IFIP international conference on dependable systems and networks workshops, Portland, OR, USA, 2019*. New York: IEEE/IFIP, pp. 9–14.

Azevedo, Rui; Iberia Medeiros; and Alysson Bessani (2019). PURE: generating quality threat intelligence by clustering and correlating OSINT. *TrustCom/BigDataSE: 2019 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering, Rotorua, New Zealand, 2019*. New York: IEEE, pp. 483–490.

Backfried, Gerhard; Christian Schmidt; Mark Pfeiffer; G. Quirchmayr; M. Glanzer; and K. Rainer (2012). Open source intelligence in disaster management. *EISIC: 2012 European intelligence and security informatics conference, Odense, Denmark, 2012*. New York: IEEE, pp. 254–258.

Bansal, Gagan; Besmira Nushi; Ece Kamar; Daniel S Weld; Walter S Lasecki; and Eric Horvitz (2019). Updates in Human-AI Teams: understanding and addressing the performance/compatibility tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 2429–2437.

Behzadan, Vahid; Carlos Aguirre; Avishek Bose; and William Hsu (2018). Corpus and deep learning classifier for collection of cyber threat indicators in twitter stream. *Big data: 2018 IEEE international conference on big data, Seattle, Washington, USA, 2018*. New York: IEEE, pp. 5002–5007.

Bernard, Rose; Gemma Bowsher; Ciaran Milner; Peter Boyle; Preeti Patel; and Richard Sullivan (2018). Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks. *Journal of Public Health*, vol. 26, no. 5, pp. 509–514.

Brennan, Robert L.; and Dale J. Prediger (1981). Coefficient Kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, vol. 41, no. 3, pp. 687–699.

Burrell, Jenna (2016). How the machine ‘thinks’: understanding opacity in machine learning algorithms. *Big Data & Society*, vol. 3, no. 1, pp. 1–12.

Büscher, Monika; Catherine Easton; Charalampia Kerasidou; Maria Aléjandra Lujan; Hayley Alter Escalante; Katrina Petersen; Marie-Christine Bonnamour; David Lund; Andreas Baur; Regina Ammicht et al. (2018). The isitethical? Exchange responsible research and innovation for disaster risk management. *ISCRAM’18: Proceedings of the 15th ISCRAM Conference, Rochester, NY, USA, 2018*. Brussels: ISCRAM Association, pp. 254–267.

Büscher, Monika; Sarah Jane Becklake; Catherine Rachel Easton; Charalampa Xaroula Kerasidou; Rachel Sarah Oliphant; Katrina Gooding Petersen; Lina Jasmontaite; and Olivier Paterour (2016). ELSI Guidelines for networked collaboration and information exchange in PPDR and risk governance. *ISCRAM’16: proceedings of the ISCRAM 2016 conference, Rio de Janeiro, Brazil, 2016*. Brussels: ISCRAM, pp. 1–12.

Casanovas, Pompeu (2014). Open source intelligence, open social intelligence and privacy by design. *ECSI-2014: Proceedings of the European conference on social intelligence, Barcelona, Spain, 2014*. Aachen: CEUR Workshop Proceedings, pp. 174–185.

Casanovas, Pompeu (2017). Cyber warfare and organised crime. A regulatory model and meta-model for open source intelligence (OSINT). In M. Taddeo, and L. Glorioso (eds.): *Ethics and policies for cyber operations*. Cham: Springer, pp. 139–167.

Casanovas, Pompeu; Juan Arraiza; Felipe Melero; Jorge González-Conejero; Gila Molcho; and Montse Cuadros (2014). Fighting organized crime through open source intelligence: regulatory strategies of the CAPER Project. *JURIX 2014: The 27th international conference on legal knowledge and information systems, Krakow, Poland, 2014*. Amsterdam: IOS Press BV, pp. 189–198.

Chitkara, Abhi; Deepti Singh; Ashish Gupta; and Gaurav Varshney (2020). Intellispect: personal information search tool. *ICOIN: 2020 international conference on information networking, Barcelona, Spain, 2020*. New York: IEEE, pp. 556–561.

Chouldechova, Alexandra; Diana Benavides-Prado; Oleksandr Fialko; and Rhema Vaithianathan (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of machine learning research: conference on fairness, accountability and transparency, New York, USA, 2018*, pp. 134–148.

Christen, Markus; Bert Gordijn; Karsten Weber; Ibo van de Poel; and Emad Yaghmaei (2017). A review, of value-conflicts in cybersecurity. *The ORBIT Journal*, vol. 1, no. 1, pp. 1–19.

Cobb, Camille; Ted McCarthy; Annuska Perkins; Ankitha Bharadwaj; Jared Comis; Brian Do; and Kate Starbird (2014). Designing for the deluge: understanding & supporting the distributed, collaborative work of crisis volunteers. *CSCW'14: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing, Baltimore, Maryland, USA, 2014*. New York: ACM, pp. 888–899.

Cousins, Karlene; Hemang Subramanian; and Pouyan Esmaeilzadeh (2019). A value-sensitive design perspective of cryptocurrencies: a research agenda. *Communications of the association for information systems*, vol. 45, no. 1, pp. 511–547.

Cuijpers, Colette (2013). Legal aspects of open source intelligence – results, of the VIRTUOSO project. *Computer Law & Security Review*, vol. 29, no. 6, pp. 642–653.

Dalton, Adam; Bonnie Dorr; Leon Liang; and Kristy Hollingshead (2017). Improving cyber-attack predictions through information foraging. *Big data: 2017 IEEE international conference on big data, Boston, MA, USA, 2017*. New York: IEEE, pp. 4642–4647.

Davis, Janet; and Lisa P. Nathan (2015). Value sensitive design: applications, adaptations, and critiques. In J. van den Hoven, P.E. Vermaas, and I. van de Poel (eds.): *Handbook of ethics, values, and technological design, sources, theory, values and application domains*. Dordrecht: Springer, pp. 11–40.

Dionisio, Nuno; Fernando Alves; Pedro M. Ferreira; and Alysson Bessani (2019). Cyberthreat detection from twitter using deep neural networks. *IJCNN: 2019 international joint conference on neural networks, Budapest, Hungary, 2019*. New York: IEEE, pp. 1–8.

Domingo-Ferrer, Josep; and Alberto Blanco-Justicia (2020). Ethical value-centric cybersecurity: a methodology based on a value graph. *Science and Engineering Ethics*, vol. 26, no. 3, pp. 1267–1285.

Dzindolet, Mary T.; Scott A. Peterson; Regina A. Pomranky; Linda G. Pierce; and H.P. Beck (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 697–718.

Edwards, Matthew; Robert Larson; Benjamin Green; Awais Rashid; and Alistair Baron (2017). Panning for gold: automatically, analysing online social engineering attack surfaces. *Computers & Security*, vol. 69, pp. 18–34.

Fallmann, Hanno; Gilbert Wondracek; and Christian Platzer (2010). Covertly probing underground economy marketplaces. *DIMVA'10: International conference on detection of intrusions and malware, and vulnerability assessment, Bonn, Germany, 2010*. Berlin: Springer, pp. 101–110.

Feng, Shi; and Jordan Boyd-Graber (2019). What can AI do for me? evaluating machine learning interpretations in cooperative play. *IUI'19: Proceedings of the 24th international conference on intelligent user interfaces, Marina del Rey, CA, USA, 2019*. New York: ACM, pp. 229–239.

Franke, Ulrik; and Joel Brynielsson (2014). Cyber situational awareness - a systematic review of the literature. *Computers & Security*, vol. 46, pp. 18–31.

Friedman, Batya (1996). Value-Sensitive, Design. *Interactions*, vol. 3, no. 6, pp. 17–23.

Friedman, Batya; David G. Hendry; and Alan Borning (2017). A survey, of value sensitive design methods. *Foundations and Trends® in Human-Computer Interaction*, vol. 11, no. 2, pp. 63–125.

Friedman, Batya; Peter H. Kahn; Alan Borning; and Alina Huldtgren (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiers, I. van de Poel, and M.E. Gorman (eds.): *Early engagement and new technologies: Opening up the laboratory*. Dordrecht: Springer, pp. 55–95.

Friedman, Batya; and David G. Hendry (2019). *Value sensitive design: Shaping technology with moral imagination*. Cambridge: MIT Press.

Friedman, Batya; and Peter Kahn (2002). Human values, ethics, and design. In J. A. Jacko and A. Sears (eds.): *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. Broadway Hillsdale: L. Erlbaum Associates Inc., pp. 1177–1201.

Ghazi, Yumna; Zahid Anwar; Rafia Mumtaz; Shahzad Saleem; and Ali Tahir (2018). A supervised machine learning based approach for automatically extracting high-level threat intelligence from unstructured sources. *FIT'18: 2018 international conference on frontiers of information technology, Islamabad, Pakistan, 2018*. New York: IEEE, pp. 129–134.

Glassman, Michael; and Min Ju Kang (2012). Intelligence in the internet age: The emergence and evolution of open source intelligence (OSINT). *Computers in Human Behavior*, vol. 28, no. 2, pp. 673–682.

Gläser, Jochen; and Grit Laudel (2010). *Experteninterviews und qualitative Inhaltsanalyse: als Instrumente rekonstruierender Untersuchungen*. Wiesbaden: VS Verlag für Sozialwissenschaften, 4th edition.

Gong, Seonghyeon; Jaeik Cho; and Changhoon Lee (2018). A Reliability Comparison Method for OSINT Validity Analysis. *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5428–5435.

González-Granadillo, Gustavo; Mario Faiella; Iberia Medeiros; Rui Azevedo; and Susana Gonzalez-Zarzosa (2019). Enhancing information sharing and visualization capabilities in security data analytic platforms. *DSN-w'19: 49th annual IEEE/IFIP international conference on dependable systems and networks workshops, Portland, OR, USA, 2019*. New York: IEEE, pp. 1–8.

González-Granadillo, Gustavo; Mario Faiella; Ibéria Medeiros; Rui Azevedo; and Susana González-Zarzosa (2021). ETIP: an enriched threat intelligence platform for improving OSINT correlation, analysis, visualization and sharing capabilities. *Journal of Information Security and Applications*, vol. 58, pp. 1–15.

Hayes, Darren R.; and Francesco Cappa (2018). Open-source intelligence for risk assessment. *Business Horizons*, vol. 61, no. 5, pp. 689–697.

Hayes, Paul; Ibo van de Poel; and Marc Steen (2020). Algorithms and values in justice and security. *AI and Society*, vol. 35, no. 3, pp. 533–555.

Heath, Christian; and Paul Luff (1992). Collaboration and control. Crisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)*, vol. 1, no. 1-2, pp. 69–94.

Huang, Shin-Ying; Yen-Wen Huang; and Ching-Hao Mao (2019). A multi-channel cybersecurity news and threat intelligent engine - SecBuzzer. *ASONAM'19: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, Canada, 2019*. New York: ACM, pp. 691–695.

Huang, Shin-Ying; and Tao Ban (2019). A Topic-Based Unsupervised Learning Approach for Online Underground Market Exploration. *TrustCom/BigdataSE: 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering, Rotorua, New Zealand, 2019*. New York: IEEE, pp. 208–215.

Jo, Hyeonseong; Jinwoo Kim; Phillip Porras; Vinod Yegneswaran; and Seungwon Shin (2021). GapFinder: Finding Inconsistency of Security Information From Unstructured Text. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 86–99.

Jones, Ridley; Michael W. Beach; Melinda McClure Haughey; Will Sutherland; and Charlotte P. Lee (2021). Construction of Shared Situational Awareness in Traffic Management. *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–27.

Kallio, Hanna; Anna-Maija Pietilä; Martin Johnson; and Mari Kangasniemi (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, vol. 72, no. 12, pp. 2954–2965.

Kandias, Miltiadis; Dimitris Gritzalis; Vasilis Stavrou; and Kostas Nikoloulis (2017). Stress level detection via OSN, usage pattern and chronicity analysis: An OSINT threat intelligence module. *Computers & Security,* vol. 69, pp. 3–17.

Kandias, Miltiadis; Lilian Mitrou; Vasilis Stavrou; and Dimitris Gritzalis (2013a). YouTube user and usage profiling: Stories of political horror and security success. *ICETE'13: International Conference on E-Business and Telecommunications, Reykjavik, Iceland, 2013*. Berlin, Heidelberg: Springer, pp. 270–289.

Kandias, Miltiadis; Vasilis Stavrou; Nick Bozovic; Lilian Mitrou; and Dimitris Gritzalis (2013b). Can We Trust This User? Predicting Insider's Attitude via YouTube Usage Profiling. *UIC-ATC'13: Proceedings of the 2013 IEEE 10th International Conference on Ubiquitous Intelligence & Computing and 2013 IEEE 10th International Conference on Autonomic & Trusted Computing, Sorrento Peninsula, Italy, 2013*. New York: IEEE, pp. 347–354.

Kandias, Miltiadis; Vasilis Stavrou; Nick Bozovic; and Dimitris Gritzalis (2013c). Proactive Insider Threat Detection Through Social Media: The YouTube case. *WPES'13: proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society, Berlin, Germany, 2013*. New York: ACM, pp. 261–266.

Kannavara, Raghudeep; Jacob Vangore; William Roberts; Marcus Lindholm; and Priti Shrivastav (2019). A Threat Intelligence Tool for the Security Development Lifecycle. *ISEC'19: Proceedings of the 12th Innovations on Software Engineering Conference, Pune, India, 2019*. New York: ACM, pp. 1–5.

Kassim, Sharifah Roziah Binti Mohd; Shujun Li; and Budi Arief (2022). How national CSIRTs leverage public data, OSINT and free tools in operational practices: An empirical study. *Cyber Security: A Peer-Reviewed Journal*, vol. 5, no. 3, pp. 251–276.

Kaufhold, Marc-André; Ali Sercan Basyurt; Kaan Eyilmez; Marc Stöttinger; and Christian Reuter (2022). Cyber Threat Observatory: Design and Evaluation of an Interactive Dashboard for Computer Emergency Response Teams. *ECIS: Proceedings of the European Conference on Information Systems, Timisuara, Romania, 18-24 June 2022*. New York: IEEE, pp. 1–17.

Kaufhold, Marc-André; Jennifer Fromm; Thea Riebe; Milad Mirbabaie; Philipp Kuehn; Ali Sercan Basyurt; Markus Bayer; Marc Stöttinger; Kaan Eyilmez; Reinhard Möller; Christoph Fuchß; Stefan Stieglitz; and Christian Reuter (2021). CYWARN: Strategy and Technology Development for Cross-Platform Cyber Situational Awareness and Actor-Specific Cyber Threat Communication. *MuC'21: Mensch und Computer 2021 - Workshopband, Ingolstadt, Germany, 2021*. Bonn: Gesellschaft für Informatik e.V.

Kaufhold, Marc André; Nicola Rupp; Christian Reuter; and Matthias Habdank (2020). Mitigating information overload in social media during conflicts and crises: design and evaluation of a cross-platform alerting system. *Behaviour and Information Technology*, vol. 39, no. 3, pp. 319–342.

Kawaguchi, Yuki; Akira Yamada; and Seiichi Ozawa (2017). AI Web-Contents Analyzer for Monitoring Underground Marketplace. *CONIP'17: Neural Information Processing, Guangzhou, China, 2017*. Cham: Springer, pp. 888–896.

Kensing, Finn; and Jeanette Blomberg (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, vol. 7, no. 3, pp. 167–185.

Khurana, Nitika; Sudip Mittal; Aritran Piplai; and Anupam Joshi (2019). Preventing Poisoning Attacks on AI Based Threat Intelligence Systems. *MLSP: 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing, Pittsburgh, PA, USA, 2019*. New York: IEEE, pp. 1–6.

Koops, Bert Jaap; Jaap Henk Hoepman; and Ronald Leenes (2013). Open-source intelligence and privacy by design. *Computer Law and Security Review*, vol. 29, no. 6, pp. 676–688.

Kossakowski, Klaus-Peter (2001). *Information Technology Incident Response Capabilities*. Books on Demand.

Krueger, Richard A.; and Mary A. Casey (2015). *Focus Group: A Practical Guide for Applied Research*. Thousand Oaks: Sage Publications, 5th edition.

Kuckartz, Udo (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Basel: Beltz Juventa, 3rd edition.

Lawrence, Heather; Andrew Hughes; Robert Tonic; and Cliff Zou (2017). D-miner: A framework for mining, searching, visualizing, and alerting on darknet events. *CNS: 2017 IEEE Conference on Communications and Network Security, Las Vegas, NV, USA, 2017*. New York: IEEE, pp. 1–9.

Layton, Robert (2016). Relative Cyberattack Attribution. In R. Layton and P. A. Watters (eds.): *Automating Open Source Intelligence. Algorithms for OSINT*. Waltham: Syngress, pp. 37–60.

Levy, Yair; and Timothy J. Ellis (2006). A Systems, Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science Journal*, vol. 9, pp. 181–212.

Ley, Benedikt; Thomas Ludwig; Volkmar Pipek; Dave Randall; Christian Reuter; and Torben Wiedenhoefer (2014). Information Expertise Sharing in Inter-Organizational Crisis Management. *Computer Supported Cooperative Work (CSCW)*, vol. 23, no. 4-6, pp. 347–387.

Le Dantec, Christopher A.; Erika Shehan Poole; and Susan P. Wyche (2009). Values as lived experience: evolving value sensitive design in support of value discovery. *CHI'09:*

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 2009. New York: ACM, pp. 1141–1150.

Liao, Xiaojing; Kan Yuan; XiaoFeng Wang; Zhou Li; Luyi Xing; and Raheem Beyah (2016). Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence. *CCS'16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 2016*. New York: ACM, pp. 755–766.

Liegl, Michael; Alexander Boden; Monika Büscher; Rachel Oliphant; and Xaroula Kerasidou (2016). Designing for ethical innovation: A case study on ELSI co-design in emergency. *International Journal of Human Computer Studies*, vol. 95, pp. 80–95.

Liu, Ruyue; Ziping Zhao; Chengjun Sun; Xiaoyu Yang; Xiaoli Gong; and Jin Zhang (2017). A Research and Analysis Method of Open Source Threat Intelligence Data. *ICPCSEE'17: Data Science. Third International Conference of Pioneering Computer Scientists, Engineers and Educators, Changsha, China, 2017*. Singapore: Springer, pp. 352–363.

Liu, Sophia B. (2014). Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain. *Computer Supported Cooperative Work (CSCW)*, vol. 23, no. 4-6, pp. 389–443.

Longo, Luca; Randy Goebel; Freddy Lecue; Peter Kieseberg; and Andreas Holzinger (2020). Explainable artificial intelligence: Concepts, applications, research challenges and visions. *CD-MAKE: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Dublin, Ireland, 2020*. Cham: Springer, pp. 1–16.

Lukosch, Stephan; Heide Lukosch; Dragoş Datcu; and Marina Cidota (2015). Providing Information on the Spot: Using Augmented Reality for Situational Awareness in the Security Domain. *Computer Supported Cooperative Work (CSCW)*, vol. 24, no. 6, pp. 613–664.

Manders-Huits, Noëmi (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, vol. 17, no. 2, pp. 271–287.

May, Thomas (1994). The Concept of Autonomy. *American Philosophical Quarterly*, vol. 31, no. 2, pp. 133–144.

Mtsweni, J.; Muyowa Mutemwa; and Njabulo Mkhonto (2016). Development of a cyber-threat intelligence-sharing model from big data sources. *Journal of Information Warfare*, vol. 15, no. 3, pp. 56–68.

Mueller, Marius; and Oliver Heger (2018). Health at any cost? Investigating ethical dimensions and potential conflicts of an ambulatory therapeutic assistance system through value sensitive design. *ICIS'18: Proceedings of the 39th International Conference on Information Systems, San Francisco, CA, USA, 2018*. Atlanta: Association for Information Systems, pp. 1–17.

Neil, Lorenzo; Sudip Mittal; and Anupam Joshi (2018). Mining Threat Intelligence about Open-Source Projects and Libraries from Code Repository Issues and Bug Reports. *ISI: 2018 IEEE International Conference on Intelligence and Security Informatics, Miami, FL, USA, 2018*. New York: IEEE, pp. 7–12.

Nisioti, Antonia; George Loukas; Aron Laszka; and Emmanouil Panaousis (2021). Data-Driven Decision Support for Optimizing Cyber Forensic Investigations. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2397–2412.

Nissenbaum, Helen (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press.

Normark, Maria; and Dave Randall (2005). Local expertise at an emergency call centre. *ECSCW'05: proceedings of the ninth european conference on computer-supported cooperative work, Paris, France, 2005*. Dodrecht: Springer, pp. 347–366.

Nunes, Eric; Paulo Shakarian; and Gerardo I. Simari (2018). At-risk system identification via analysis of discussions on the darkweb. *eCrime: 2018 APWG Symposium on Electronic Crime Research, San Diego, California, USA, 2018*. New York: IEEE, pp. 1–12.

Pastor-Galindo, Javier; Pantaleone Nespoli; Felix Gomez Marmol; and Gregorio Martinez Perez (2020). The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. *IEEE Access*, vol. 8, pp. 10282–10304.

Petersen, Laura; Laure Fallou; Paul Reilly; and Elisa Serafinelli (2017). Public expectations of social media use by critical infrastructure operators in crisis communication. *ISCRAM'17: Proceedings of the 14th ISCRAM Conference, Albi, France, 2017*. Brussels: ISCRAM, pp. 1–10.

Pournouri, Sina; Shahrzad Zargari; and Babak Akhgar (2019). An Investigation of Using Classification Techniques in Prediction of Type of Targets in Cyber Attacks. *ICGS3: 2019 IEEE 12th international conference on global security, safety and sustainability, London, UK, 2019*. New York: IEEE, pp. 202–212.

Pournouri, Sina; and Babak Akhgar (2015). Improving cyber situational awareness through data mining and predictive analytic techniques. *ICGS3: International Conference on Global Security, Safety, and Sustainability, London, UK, 2015, Vol. 534 of ICGS3*. Cham: Springer, pp. 21–34.

Purohit, Hemant; Andrew Hampton; Shreyansh Bhatt; Valerie L. Shalin; Amit P. Sheth; and John M. Flach (2014). Identifying seekers and suppliers in social media communities to support crisis coordination. *Computer Supported Cooperative Work (CSCW)*, vol. 23, no. 4, pp. 513–545.

Quick, Darren; and Kim-Kwang Raymond Choo (2018). Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+OSINT): A timely and cohesive mix. *Future Generation Computer Systems*, vol. 78, pp. 558–567.

Rajamäki, Jyri (2019). Design Science Research Towards Privacy by Design in Maritime Surveillance ICT Systems. *Information & Security: An International Journal*, vol. 43, no. 2, pp. 196–214.

Rajamäki, Jyri; and Jussi Simola (2019). How to apply privacy by design in OSINT and big data analytics. *ECCWS'19: Proceedings of the 18th European Conference on Cyber Warfare and Security, Coimbra, Portugal, 2019*. Reading: Academic Conferences and Publishers International, pp. 364–371.

Ranade, Priyanka; Sudip Mittal; Anupam Joshi; and Karuna Joshi (2018). Using Deep Neural Networks to Translate Multi-lingual Threat Intelligence. *ISI: 2018 IEEE International Conference on Intelligence and Security Informatics, Miami, FL, USA, 2018*. New York: IEEE, pp. 238–243.

Randall, David; Richard Harper; and Mark Rouncefield (2007). *Fieldwork for design: theory and practice*. London: Springer Science & Business Media.

Reuter, Christian; Marc-André Kaufhold; Thomas Spielhofer; and Anna Sophie Hahne (2017). Social media in emergencies: A representative study on citizens' perception in germany. *Proceedings of the ACM on human-computer interaction*, vol. 1, no CSCW, pp. 1–19.

Reuter, Christian; Thomas Ludwig; and Volkmar Pipek (2014). Ad Hoc Participation in Situation Assessment: Supporting Mobile Collaboration in Emergencies. *ACM Transactions on Computer-Human Interaction*, vol. 21, no. 5, pp. 1–26.

Riebe, Thea; Marc-André Kaufhold; and Christian Reuter (2021a). The Impact of Organizational Structure and Technology Use on Collaborative Practices in Computer Emergency Response Teams: An Empirical Study. *Proceedings of the ACM on human-computer interaction*, vol. 5, no. CSCW2, pp. 1–30.

Riebe, Thea; Tristan Wirth; Markus Bayer; Philipp Kühn; Marc-André Kaufhold; Volker Knauthe; Stefan Guthe; and Christian Reuter (2021b). CySecAlert: An Alert Generation System for Cyber Security Events Using Open Source Intelligence Data. *ICICS: International Conference on Information and Communications Security, Chongqing, China, 2021*, pp. 429–446.

Robinson, Sandra L. (1996). Trust and Breach of the Psychological Contract. *Administrative Science Quarterly*, vol. 41, no. 4, pp. 574–599.

Rädiker, Stefan; and Udo Kuckartz (2019). *Analyse qualitativer Daten mit MAXQDA*. Wiesbaden: Springer.

Schaffer, James; John O'Donovan; James Michaelis; Adrienne Raglin; and Tobias Höllerer (2019). I can do better than your AI: expertise and explanations. *IUI '19: proceedings of the 24th international conference on intelligent user interfaces, Marina del Rey, CA, USA, 2019.* New York: ACM, pp. 240–251.

Schäfer, Matthias; Markus Fuchs; Martin Strohmeier; Markus Engel; Marc Liechti; and Vincent Lenders (2019). BlackWidow: Monitoring the Dark Web for Cyber Security Information. *CyCon: 2019 11th International Conference on Cyber Conflict: Silent Battle, Tallinn, Estonia, 2019.* Tallinn: NATO CCD COE, pp. 1–21.

Simran, K.; Prathiksha Balakrishna; R. Vinayakumar; and K. P. Soman (2020). Deep Learning Approach for Enhanced Cyber Threat Indicators in Twitter Stream. In S. M. Thampi, G. Martinez Perez, R. Ko, and D. B. Rawat (eds.): *Security in Computing and Communications*, Vol. 1208. Singapore: Springer Singapore, pp. 135–145. Series Title: Communications in Computer and Information Science.

Skopik, Florian; Giuseppe Settanni; and Roman Fiedler (2016). A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, vol. 60, pp. 154–176.

Soden, Robert; and Leysia Palen (2018). Informating crisis: Expanding critical perspectives in crisis informatics. *Proceedings of the ACM on Human-Computer interaction*, vol. 2, no. CSCW, pp. 1–22.

Syed, Romilla (2020). Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system. *Information & Management*, vol. 57, no. 6, p. 103334.

Tapia, Andrea H.; and Kathleen Moore (2014). Good enough is good enough: Overcoming disaster response organizations' slow social media data adoption. *Computer Supported Cooperative Work (CSCW)*, vol. 23, no. 4-6, pp. 483–512.

Tundis, Andrea; Samuel Ruppert; and Max Mühlhäuser (2020). On the Automated Assessment of Open-Source Cyber Threat Intelligence Sources. *ICCS: International Conference on Computational Science, Amsterdam, The Netherlands, 2020*. Cham: Springer, pp. 453–467.

Turilli, Matteo; and Luciano Floridi (2009). The ethics of information transparency. *Ethics and Information Technology*, vol. 11, no. 2, pp. 105–112.

Urban, Tobias; Matteo Große-Kampmann; Dennis Tatang; Thorsten Holz; and Norbert Pohlmann (2020). Plenty of Phish in the Sea: Analyzing Potential Pre-Attack Surfaces. *ESORICS'20: European Symposium on Research in Computer Security, Guildford, UK, 2020*. Cham: Springer, pp. 272–291.

Vacas, Ivo; Iberia Medeiros; and Nuno Neves (2018). Detecting Network Threats using OSINT Knowledge-Based IDS. *EDCC: 2018 14th European Dependable Computing Conference, Iaşi, Romania, 2018*. New York: IEEE, pp. 128–135.

Van der Kleij, Rick; Geert Kleinhuis; and Heather Young (2017). Computer security incident response team effectiveness: A needs assessment. *Frontiers in Psychology*, vol. 8, p. 2179.

Wang, Danding; Qian Yang; Ashraf Abdul; and Brian Y. Lim (2019). Designing theory-driven user-centric explainable AI. *CHI'19: Proceedings of the 2019 CHI conference on human factors in computing systems, Glasgow, Scotland, UK, 2019*. New York: ACM, pp. 1–15.

Westin, Alan F. (1967). *Privacy and Freedom*. London: The Bodley Head.

Williams, Heather; and Ilana Blum (2018). *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. RAND Corporation.

Wright, David; and Michael Friedewald (2013). Integrating privacy and ethical impact assessments. *Science and Public Policy*, vol. 40, no. 6, pp. 755–766.

Wulf, Volker; Markus Rohde; Volkmar Pipek; and Gunnar Stevens (2011). Engaging with Practices: Design Case Studies as a Research Framework in CSCW. *CSCW'11: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, Hangzhou, China, 2011*. New York: ACM, pp. 505–512.

Zhang, Qiaoning; Matthew L. Lee; and Scott Carter (2022). You Complete Me: Human-AI Teams and Complementary Expertise. *CHI'22: Conference on Human Factors in Computing Systems, New Orleans, LA, 2022*. New York: ACM, pp. 1–28.

van de Poel, Ibo (2011). The Relation Between Forward-Looking and Backward-Looking Responsibility. In N.A. Vincent, I. van de Poel, and J. van den Hoven (eds.): *Moral responsibility: Beyond free will and determinism*. Dordrecht: Springer Netherlands, pp. 37–52.

van de Poel, Ibo (2020). Core Values and Value Conflicts in Cybersecurity: Beyond Privacy Versus Security. In M. Christen, B. Gordijn, and M. Loi (eds.): *The Ethics of Cybersecurity*. Cham: Springer, pp. 45–71.

van de Poel, Ibo; and Lambér Royakkers (2011). *Ethics, Technology, and Engineering: An Introduction*. Malden: Wiley-Blackwell.

vom Brocke, Jan; Alexander Simons; Kai Riemer; Björn Niehaves; Ralf Plattfaut; and Anne Cleven (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *Communications of the Association for Information Systems*, vol. 37, no. 1, pp. 205–224.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.