Artificial Intelligence for Cybersecurity: Towards Taxonomy-based Archetypes and Decision Support

Jana Gerlach
Information Systems Institute, Leibniz Universität Hannover, gerlach@iwi.uni-hannover.de

Oliver Werth
Leibniz Universität Hannover, werth@iwi.uni-hannover.de

Michael H. Breitner
Leibniz Universität Hannover, breitner@iwi.uni-hannover.de

Artificial Intelligence for Cybersecurity: Towards Taxonomy-based Archetypes and Decision Support

Completed Research Paper

Jana Gerlach
Leibniz University Hannover
Königsworther Platz 1
30167 Hanover, Germany
gerlach@iwi.uni-hannover.de

Oliver Werth
Leibniz University Hannover
Königsworther Platz 1
30167 Hanover, Germany
werth@iwi.uni-hannover.de

Michael H. Breitner
Leibniz University Hannover
Königsworther Platz 1
30167 Hanover, Germany
breitner@iwi.uni-hannover.de

Abstract

Cybersecurity is a critical success factor for more resilient companies, organizations, and societies against cyberattacks. Artificial intelligence (AI)-driven cybersecurity solutions have the ability to detect and respond to cyber threats and attacks and other malicious activities. For this purpose, the most important resource is security-relevant data from networks, cloud systems, clients, e-mails, and previous cyberattacks. AI, the key technology, can automatically detect, for example, anomalies and malicious behavior. Consequently, the market for AI-driven cybersecurity solutions is growing significantly. We develop a taxonomy of AI-driven cybersecurity business models by classifying 229 real-world services. Building on that, we derive four specific archetypes using a cluster analysis toward a comprehensive academic knowledge base of business model elements. To reduce complexity and simplify the results of the taxonomy and archetypes, we propose DETRAICS, a decision tree for AI-driven cybersecurity services. Practitioners, decision-makers, and researchers benefit from DETRAICS to select the most suitable AI-driven service.

Keywords: Artificial intelligence, AI-driven cybersecurity, taxonomy, archetypes, decision tree

Introduction

Cyberspace is exposed to a variety of risks resulting from physical and cyber threats amplified by the progressing digital transformation among the most affecting technologies such as the internet of things, digital platforms, cloud computing, and artificial intelligence (AI) (CISA 2022; European Parliament 2022a). A growing volume, velocity, and variety of data is collected, processed, and transmitted in cyberspace which is by design interconnected with data from the digital and physical world, thus emerging new dangers in the form of cyberattacks (CISA 2022; ENISA 2022). According to the European Parliament (2022b), cyberattacks are "attempts to misuse information, by stealing, destroying or exposing it, aiming to disrupt or destroy computer systems and networks." The main cybersecurity threats in 2021 were ransomware, data breaches - and leaks, malware, disinformation, human errors, threats against availability and integrity, as well as e-mail-related and supply chain threats. Especially, sectors of critical infrastructure such as health, energy, transport, and finance are targeted by attackers. At the moment, ransomware is considered the most concerning threat as, for example, every 11 seconds, a corporate ransomware attack occurs (European Parliament 2022b). The costs of cybercrime involve significant monetary consequences as well as the impact of data disclosure, reputational damage, and loss of trust from customers and business partners (Sarker et al. 2021). Mitigating these cyber threats requires cybersecurity measures that exceed the capabilities and skills of cyber-attackers. Cybersecurity refers to the utilization of management strategies, practices, and technologies, to ensure the safety of data, computers, networks, and programs from malicious activities (European Parliament 2022b). Traditional security measures include antivirus programs, encryption, firewalls, and authentication of users, but they cannot fulfill today's security requirements, which addresses the need for more advanced technologies for cybersecurity (Sarker et al. 2021). Thus, makes an interesting field to investigate.

Besides the increased risk of the growing amount of data in cyberspace, these data, in turn, have the potential to exploit security-relevant information to protect against malicious actors. AI techniques such as machine learning (ML), deep learning, and natural language processing can conduct, for example, detection analysis for intrusions, anomalies, and malicious behavior, as well as the classification of attacks and malware traffic (Sarker et al. 2021). Therefore, AI-driven cybersecurity measures allow the self-learning analysis of security-relevant information and identification of behavioral patterns from networks, cloud systems, clients, e-mails, and previous cyberattacks. The market for AI-driven cybersecurity services expanded in the last years, as the worldwide market value rose from 8.8 billion US-Dollars in 2019 to 10.5 billion US-Dollars in 2020 and is forecasted to increase to 46.3 billion US-Dollars by 2027 (Pillsbury 2021; Statista 2022). The rising market value shows promising chances for responsible stakeholders, such as cybersecurity managers or venture capitalists to invest money into new market entrants. However, the current state of academic literature does not provide a comprehensive and empirically-validated analysis of real-world AI-driven cybersecurity services. As a result, past researchers call for the ignition of such carefully-conducted research about this market in general (Wallace et al. 2020) and a specific look at AI-driven cybersecurity solutions (Sarker et al. 2021). A classification around AI-driven cybersecurity can be advantageous since it reduces the market's ongoing complexity for academics studying crucial business model components and their interrelations. On the other hand, practitioners need a clear overview of what the alternatives in the cybersecurity market are and how to choose adequate cybersecurity services. In addition, better knowledge of the services provided can enhance the business relationships between services that offer AI-driven cybersecurity services and customers towards efficient protection against cyber risk (Croasdell and Palustre 2019). Based on our motivations, we address the research questions (RQs):

RQ1: *Which archetypes of AI-driven cybersecurity solutions can be deduced empirically from a taxonomy of corresponding business models?*

RQ2: *Which dimensions and characteristics must be integrated into a decision support framework to encourage responsible stakeholders to select an adequate and efficient AI-driven cybersecurity service?*

Our contributions are threefold: First, we develop a taxonomy of AI-driven cybersecurity business models and services following the methodology for taxonomy development by Nickerson et al. (2013) and Kundisch et al. (2021). Using the "most prominent and widely used approach in the field" (Schöbel et al. 2020: 647), we can build a comprehensive knowledge foundation of differences and similarities in the area of interest. Taxonomies are a valuable outcome to examine those differences (Weking et al. 2020). Also, they can be used as a starting point for theory-building, like design theories, towards a better understanding of AI-driven business models and their services (Muntermann et al. 2015; Kundisch et al. 2021). Second, to evaluate the proposed taxonomy, we use clustering techniques, as Kundisch et al. (2021) suggested. With this evaluation, we identify meaningful specific archetypes between the services provided and check the applicability of taxonomies information. Furthermore, clustering techniques are advantageous since they can provide groups of objects instead of individual objects to reduce complexity and go beyond the descriptive nature of taxonomies. Although taxonomies and archetypes can be overwhelming for practitioners, we third propose *DETRAICS,* a **de**cision **tr**ee for **AI**-driven **c**ybersecurity **s**ervices. Decision trees can help practitioners to reduce uncertainties and present all alternatives to a specific problem (e.g., Magee 1964). We evaluate the taxonomy, the clusters, and *DETRAICS* with expert interviews. Practitioners and decision-makers can use *DETRAICS* to choose the best AI-driven service suitable for them. Also, they are an intuitive complementary to already existing checklists for cybersecurity decision-making as proposed, such as by Wallace et al. (2020). Furthermore, start-ups and already existing companies can use our taxonomy and archetypes to evaluate what the market is doing and tailor their provided offerings.

The paper is structured as follows: We provide a theoretical background on AI and AI-driven cybersecurity applications. Afterward, we describe our research design, research methods, and data collection. Subsequentially, we deduce our taxonomy, archetypes, and our *DETRAICS* decision tree. We discuss our results and findings with insights gained by evaluation interviews with various stakeholders. Implications and recommendations for academics and practitioners are presented. We conclude with limitations, further research ideas, and conclusions.

Theoretical Background

Artificial Intelligence Foundations

AI is a technology with a unique capability to learn, sense, reason, solve problems, act, interpret language, and plan (Kumar 2017; Jöhnk et al. 2021). It has numerous different application possibilities and can be defined as "the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system's ultimate goal" (Albus 1991:474). Due to the exponential increase in data volume, variety, and velocity, AI has advanced from theory to real-world applications and can be utilized in various sectors, such as medicine, security, and energy (Babatunde et al. 2020). In general, AI has two main goals: The first one is the production and development of AI technologies that have the ability to solve real-world problems in different areas. The second one is to use scientific modeling approaches and algorithms to ensure that data can be processed and analyzed scientifically. This is an aspect unseen by humans and is done by analyzing and processing the data using various models and algorithms. As a result, business intelligence capabilities are enhanced since it can process huge datasets (so-called Big Data), extract patterns, and design graphs for business intelligence (Kumar 2017). A subset of AI is ML. It exhibits all the experiential learning factors of human intelligence and its capabilities of learning and improving its analysis through computational algorithms. ML is a special subset of AI since machines can be programmed to learn from data, which makes it the most promising tool in the AI toolbox for business today. Therefore, in contrast to traditional programming, ML allows a computer program to learn to recognize patterns on its own and predict what may happen based on its discovery (Padmanabhan et al. 2022). Another important term in this field of AI is deep learning, a subset of ML. Deep learning techniques are used to solve real-world problems by utilizing neural networks to emulate human decision-making. The problem is that deep learning can be very expensive, as massive datasets are required for the training. This is primarily since a learning algorithm has to consider a large number of parameters, and this can initially produce many false positives, for example, instruct a deep learning algorithm to learn how an animal looks. A very large dataset is required to grasp the smallest details (LeCun et al. 2015). AI and ML techniques and tools are being used and applied to solve real-world problems and provide solutions. These application areas are natural language processing, computer vision, predictive analytics, and robotics. These technologies can be further extended and used to satisfy industrial, e.g., cybersecurity needs (Kumar et al. 2020).

Artificial Intelligence Applications for Cybersecurity and Business Models

Cybersecurity refers to the protection of the confidentiality, integrity, and availability of computer systems, networks, data, and information. These three main objectives of cybersecurity are called the CIA triad. Confidentiality means that information, files, usernames, passwords, etc., can only be accessed by authorized persons, devices, or processes. Integrity describes the protection against change or manipulation of data. Availability protection is intended to prevent hardware, software, and process failures and malfunctions (Nweke 2017). According to Aftergood et al. (2017) and Craigen et al. (2014), cybersecurity is defined as the utilization of a set of tools, technologies, practices, and processes to protect networks, data, systems, hard- and software from attacks, damages, and intrusions. The most common cybersecurity attacks and threats are unauthorized access (Sun et al. 2018), ransomware (McIntosh et al. 2019), malware, denial of service, social engineering (Jang-Jaccard and Nepal 2014), phishing (Alsayed and Bilgrami 2017), insider threats (Warkentin and Willison 2009), data breaches (Shaw 2009), and supply chain attacks (Ohm et al. 2020). Such security incidents are rooted in external and internal intrusions and have the potential to cause significant damage (Xin et al. 2018; Sarker et al. 2021). Internal intrusion happens from inside the company by authorized users identified as insider threats. This is done, for example, by employees to abuse network and asset access (Sun et al. 2018; Sarker et al. 2021). External intrusion is caused by unauthorized access to networks, systems, data, malicious software, and (multiple) infected internet-connected devices performed by cybercriminals (Sarker et al. 2021). Because of their computing power and capabilities, AI and ML have the potential to play a significant role in the cybersecurity domain. These tasks and approaches can be, for example, but are not limited to the utilization of detecting intrusions through clustering techniques (Sharifi et al. 2015). Another example is the identification of malicious activities, attacks, and anomalies in networks and systems (Moon et al. 2017) and classifying attacks and malware traffic by monitoring and analyzing behavior and activities in network databases, users, and applications (Yin et al. 2017). Furthermore, AI can help to prevent cyberterrorism (Hansen et al. 2007). The used AI techniques are diverse and include clustering, random forest, support vector machines, neuronal networks, and deep learning (Xin et al. 2018). While several academic articles on cybersecurity with an application of AI exist from a technical perspective or a specific look with case studies, a holistic institutional perspective is somewhat overlooked by academics. As a result, the literature on possible AI-driven business model elements is fragmented and empirically validated, and comprehensive research is still missing. AI shows emerging capabilities and possibilities for cybersecurity services. Furthermore, as described in the introduction section, the market value rises and is forecasted to increase. Therefore, it can be assumed that the market will become attractive for new market entrants who want to challenge incumbent cybersecurity firms. Two of these incumbents are Darktrace and Cowbell Cyber. Darktrace is an AI company, located in the United Kingdom and founded in 2013 that offers cybersecurity solutions to identify and eliminate insider threats, zero-day malware, data loss, supply chain risk, and industrial espionage. The applied self-learning AI framework is modeled on the human immune system that reacts in real-time and responds with measures autonomously (Darktrace 2022). In contrast, Cowbell Cyber provides customized cyber insurance for enterprises, is located in the United States, and was founded in 2019. To evaluate the clients' individual cyber risk exposure and coverage selection, Cowbell Cyber applies an AI-based model that continuously monitors the company's status of cyber risk mitigation, recovery, and response (Cowbell Cyber 2022). Darktrace and Cowbell Cyber were very successful with the provided services and showed fruitful market survival. However, the market will become competitive with new market entrants, making it difficult to overlook the services provided by interested stakeholders. Therefore, we argue that meaningful support for decision-makers to choose the most suitable AI-driven cybersecurity solution for their purposes is a valuable outcome and useful in cybersecurity practice. In addition, existing players in the cybersecurity market can take advantage of our current overview, gain insight into alternatives, and adapt their services accordingly.

Research Design, Research Methods, and Data Collection

To answer our RQs, we followed a four-stage research design. We first developed a taxonomy, followed by clustering the taxonomic results (stage 2). We deduced a decision tree from our dimensions, characteristics, and archetypes derived from the two stages before. Taxonomies contribute to theory building and practice frameworks as they strengthen the rigor understanding and allow to identify novel design options of, for instance, business models. The identification of archetypes enables a generalization of the empirically analyzed services and a distinction between business model types. This enhances the knowledge base as well as the perception, assembling, and innovation of business models and their value creation (Möller et al. 2021). Besides the contribution of the taxonomy and archetypes to theory and practice, they can also be further utilized providing a basis for decision tree development. The complementary visualization provided by a decision tree based on our taxonomy's and our archetypes' information improves the comprehensibility and applicability of the findings (Mueller et al. 2022). In the fourth stage, we evaluated our results and findings with expert interviews. Our overall research design is summarized in Table 1.

In the first step of our research design, we developed a taxonomy of real-world AI-driven cybersecurity services according to Nickerson et al. (2013) and Kundisch et al. (2021). Taxonomies are a valuable tool since they can help structure or organize a domain of interest, like business models or services. They help to grasp possible similarities or differences between objects (Szopinski et al. 2019). Typically, the taxonomic development begins with the definition of the ending conditions as well as the meta-characteristic. The meta-characteristic is the most inclusive characteristic that serves as the basis for all dimensions and characteristics that follow in the process (Nickerson et al. 2013). We intended to examine the theoretically grounded and empirically validated elements of AI-driven cybersecurity business models and elements. As a result, we define our meta-characteristic as *business elements of AI-driven cybersecurity services from the perspective of its customers*. Hence, we identified the underlying business model based on the value creation process and the offering of the services.

	Stage 1: Taxonomy Development			**Stage 2: Clustering**	**Stage 3: Decision Tree Development**	**Stage 4: Evaluation of the Results**
	Step 1: Meta-characteristic, ending conditions, first conceptual-to-empirical approach	**Step 2:** Business model dataset creation	**Step 3:** Four empirical-to-conceptual approaches, one additional conceptual-to-empirical approach	**Step 4:** Taxonomy application and evaluation	**Step 5:** Expert system development	**Step 6:** Expert consultation
Tasks	1.1 Defining meta-characteristic 1.2 Defining ending conditions 1.3 Systematic keyword search in academic databases 1.4 Literature analysis 1.5 Concept matrix 1.6 First iteration C2E	2.1 Advanced company search 2.2 Keyword-based search 2.3 Search in the cybersecure. industry	3.1 Second iteration E2C 3.2 Third iteration C2E 3.3 Fourth iteration E2C 3.4 Fifth iteration E2C 3.5 Sixth iteration E2C	4.1 Identification of the optimal number of clusters 4.2 Cluster analysis	5.1 Split dataset in training and test data 5.2 Run algorithm	6.1 Determination of evaluation criteria 6.2 Conduction of expert interviews
Method / Reference	Taxonomy development (Nickerson et al. 2013; Kundish et al. 2021); Literature review (Webster and Watson 2002)	Advanced search at crunchbase. com (Crunchbase 2022)	Taxonomy development (Nickerson et al. 2013; Kundisch et al. 2021)	Cluster analysis (Kaufman and Rousseeuw 1990) via RStudio; Taxonomy evaluation (Kundisch et al. 2021)	Decision tree development (Pedregosa et al. 2011) via scikit-learn	Taxonomy evaluation with expert interviews (Kundisch et al. 2021)
Data	15 academic articles in the context of cybersecurity services	List of AI-driven cyber-security services	Preliminary taxonomy and dataset of real-world cybersecurity services, additional literature	Classified dimensions and characteristics with real-world services, i.e., the taxonomy	Classified real-word services and corresponding archetypes	Three focus group interviews with seven experts
Results and Findings	Knowledge base for conceptual-to-empirical iteration, preliminary taxonomy with meta-characteristic	Dataset for empirical-to-conceptual iteration	Taxonomy of cybersecurity service design options according to ending conditions	Archetypes of cybersecurity business models and patterns	Decision support framework for cybersecurity stakeholders	Results regarding the completeness, usefulness, and appropriateness of the taxonomy, archetypes, and *DETRAICS*

Table 1. Research Design

Before starting with the iterative taxonomy development process, we defined the ending conditions, which serve as a criterion that implies that if all ending conditions are met, the taxonomy development process can be stopped. We decided to use the seven objective, and five subjective ending conditions as Nickerson et al. (2013) proposed. Furthermore, we derived an initial set of dimensions and characteristics. Therefore, our taxonomy development process started with a conceptual-to-empirical (C2E) approach to include scientific knowledge on cybersecurity, AI, and digital business models. To identify relevant literature, we conducted a systematic keyword-based literature search in the databases: AISeL, IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and Web of Science with the search string ("cybersecurity" OR "cyber security" AND "artificial intelligence" OR "AI" OR "machine learning" OR "ML" AND "taxonomy" OR "business model" OR "classification" OR "service"). This search leads to a meaningful set of 15 scientific articles related to areas of AI technology (i.e., Xin et al. 2018; Weber et al. 2022), cyberattacks (i.e., Mullet et al. 2021), defense strategies, and cybersecurity services (i.e., Sarker et al. 2020, 2021). To systematically derive an initial set of dimensions, we developed a concept matrix as Webster and Watson (2002) proposed (see Table A in online Appendix). Based on this matrix, we used the six derived concepts (according to the terminology of Webster and Watson 2002) as six initial dimensions from the first iteration that serve as a starting point for the next iterations.

For the second step of the research procedure, we used the online company database 'crunchbase.com' (Crunchbase 2022) as our primary data source. Data collection took place in March 2022. To find the relevant services, we conducted an advanced company search and searched for the keywords ("AI" OR "artificial intelligence" OR "ML" OR "machine learning") in the cybersecurity industry. This procedure results in a list of a total of 846 services. Due to the lack of information on the respective company websites and the presentation of the services, we examine a final set of 229 services, i.e., services that serve as a basis for the forthcoming empirical-to-conceptual (E2C)-approaches.

In the third step, we performed five approaches that were iteratively conducted as proposed by Nickerson et al. (2013). At the end of each iteration, we checked and discussed if our taxonomy fulfilled the subjective and objective ending conditions. We first classify five services with the dimensions derived from the C2E-approach since we identified new dimensions and characteristics. After the analysis of 5 services, we conducted a further C2E iteration to include scientific knowledge after analyzing real-world services, which allows the focus on specific aspects. We continued the E2C iteration process until the taxonomy reached stability. This means that no new dimensions arose, and ending conditions were met. Consequently, we classified all 229 services. We identified 13 dimensions and 58 characteristics. Figure 1 shows the iterative taxonomy development process, i.e., steps 1 to 3 in our research design.

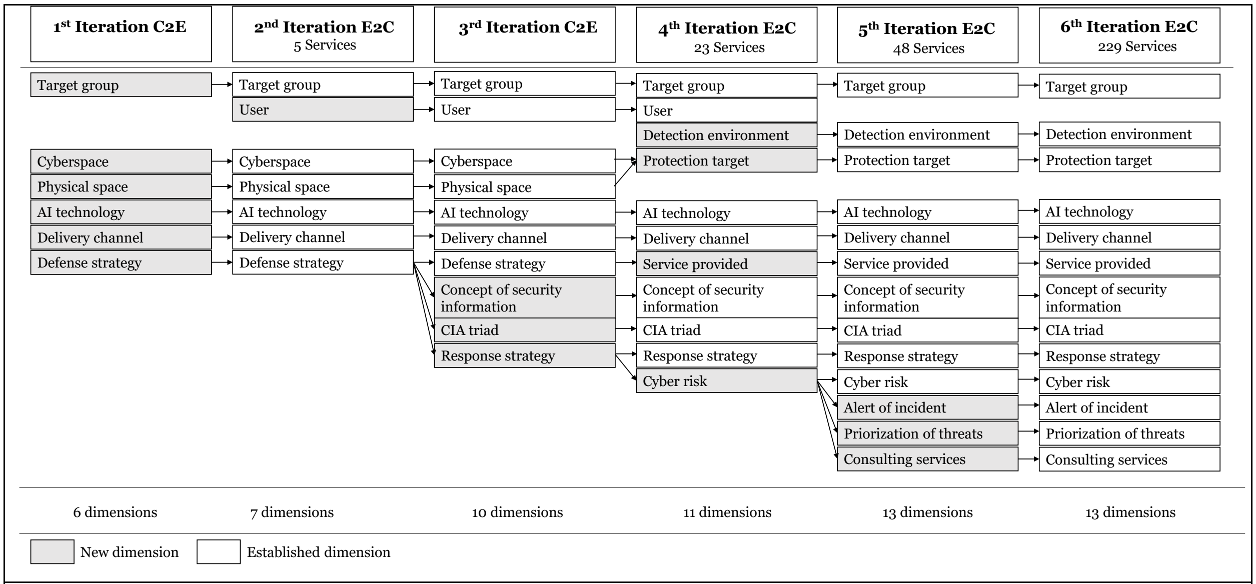

Figure 1. Iterative Taxonomy Development Process

In the fourth step, we evaluated and checked the applicability of the taxonomy by following guidelines for taxonomy designers according to Kundisch et al. (2021). Therefore, we conducted a cluster analysis with R-Studio based on the taxonomy's dimensions and characteristics to empirically identify archetypes of AI-driven cybersecurity services. In general, our cluster analysis finds groups (i.e., business models) within objects (i.e., AI-driven cybersecurity services) that minimize their differences and maximize differences between groups (Kaufman and Rousseeuw 1990). We applied the k-means cluster analysis, because it has the advantage of being easy to use with unlabeled data and large datasets. Furthermore, k-means clustering can be used in a variety of application areas and its results are easy to interpret and visualize due to the clusters formed (Punj and Stewart 1983; Likas et al. 2003). However, the number of clusters must be defined in advance. For this number identification, we applied the "Silhouette" and "Elbow" methods that visually represent and determine the quality of the created clusters by measuring their cohesion and separation (Saputra et al. 2020). The results of the "Elbow" method indicated four clusters as the optimal number of clusters (see online Appendix). The archetypes derived from the cluster analysis are useful to complement the knowledge of taxonomy provided and have the ability to go beyond their descriptive nature (Möller et al. 2021). By identifying four specific clusters, we answer RQ1.

To answer RQ2, we developed in the fifth step a decision tree that supports responsible stakeholders in selecting the most appropriate and efficient AI-driven cybersecurity solution for own purposes. The decision tree has the advantage of graphical visualization. Thus, the decision tree is easy to understand and to interpret, which reduces the threshold of entry and the degree of prior explanation needed. For our decision tree development, we followed the guidelines by Pedregosa et al. (2011) via sci-kit learn. The vectors of cybersecurity characteristics derived from real-world services' classification are the input data to the model, and the archetypes are the decision classes targeted by the algorithm for prediction. Since the taxonomy's archetypes, dimensions, and characteristics can be overwhelming and seem to be "too academic" for practitioners or responsible stakeholders, we provide an easy decision tree that guides unique suggestions. Decision trees are extremely useful for management and visualizing various options that can be considered for important decisions (Magee 1964), i.e., AI-driven cybersecurity services, here.

In the sixth step, we created an interview guideline for interviews with experts, focusing on evaluating our results and findings regarding completeness, usefulness, and appropriateness for relevant target user groups such as practitioners and consultants. We oriented ourselves on guidance for taxonomy evaluation proposed by Kundisch et al. (2021). Sub-sequentially, we interviewed seven practitioners (three focus group interviews) with domain-specific experts in the field of cybersecurity. Experts 1 and 2 are located in the energy sector and work in the cybersecurity division of a German energy provider. Experts 3 and 4 are consultants for cybersecurity, for example, in the financial services sector. Experts 5, 6, and 7 are consultants in the cybersecurity sector. All interviewees were recruited through the social networks of the authors. Interviews took place online or face-to-face and were held in August 2022. We took notes from the interviews documenting relevant statements from the interviews. These notes served as an additional knowledge foundation for the discussion and evaluation section.

Results and Findings

AI-driven Cybersecurity Business Models Taxonomy

We identified 13 dimensions and 58 characteristics among 229 services with our six iterations, see Table 2. While dimensions and characteristics that are not intuitive to the reader, they will be defined sub-sequently. Several observations can be made from this taxonomy: First, we found a clear tendency for business or business and government as the *target group* of such services. The investigated services are mainly focused on defending cyberspaces (*protection target*), including application programming interfaces, internet accesses, and (internal or external) e-mails. The minority offers services to protect physical spaces (5 services) or brands (identified only two times). *Services provided* are identified as diverse. We found services focusing on resilience (51 times) or cyberattack prevention (6), while the majority offer multiple services (92). We checked whether *AI technology* is used or named from the services' side. While 106 services did not specify their underlying AI technology to their (potential) customers, ML was most mentioned (98). Most of the investigated services deliver their AI-driven cybersecurity solution via a (cloud) platform (*delivery channel*). With *detection environment,* we evaluate which area is scanned for cyber risks. Most of the services scanned internal business environments (*detection environment*) from the (potential) customers (106), but also internal business environments and internet connections are identified often (84). We define c*yber risks* as the source of the risk coming from. Our investigation identified that most services want to protect from third-party risks (160), while seven are focused on the first party risks (internal treats). The clear majority of investigated objects provide an automated (87) or manual (116) *response strategy* against cyberattacks. At services with the manual characteristic, the customer can perform countermeasures by hand. We defined the *concept of security information* as a dimension that describes the deepness of the information that the cybersecurity service provides to its intended customers. Here, different characteristics were identified. While 105 services provide information on the threat and vulnerability, i.e., showing the system's scanned weaknesses, 35 services focus on attacks, i.e., showing attacks from criminals. We also found services that show an additional (negative) impact for the customer by an attack. Also, controls are provided. By controls, we mean the provision of possible control mechanisms against incidents that the customer can implement. According to the typology of *CIA triad*, as described in the theoretical background section, we found that the offered AI-driven cybersecurity service wants to protect multiple factors (124), while confidentiality was named as the second largest factor (63). 138 services *alert* the customer if an *incident* occurs, e.g., through short message services, e-mail, or notification through dashboard. Some services *prioritize the threats* they identify and display them to the customer (89). The minority of services (58) provide additional *consulting services*, for example, coaching for cybersecurity awareness.

Dimension D_i	**Characteristics $C_{i,j}$**	
D_1 Target group	$C_{1,1}$ Individuals (8)	$C_{1,2}$ Business (184)
	$C_{1,3}$ Governments (2)	$C_{1,4}$ Business+governments (26)
	$C_{1,5}$ Multiple (9)	
D_2 Protection target	$C_{2,1}$ Cyberspace (121)	$C_{2,2}$ Physical space (5)
	$C_{2,3}$ Cyber+physical space (67)	$C_{2,4}$ Brand protection (2)
	$C_{2,5}$ Multiple (33)	$C_{2,6}$ Others (1)
D_3 Service provided	$C_{3,1}$ Resilience (51)	$C_{3,2}$ Cyberattack prevention (6)
	$C_{3,3}$ Intrusion detection (21)	$C_{3,4}$ Cyberattack prevention+intrusion detection (30)
	$C_{3,5}$ Cyberattack prevention+response (10)	$C_{3,6}$ Intrusion detection+response (19)
	$C_{3,7}$ Multiple (92)	
D_4 AI technology	$C_{4,1}$ No specified AI (107)	$C_{4,2}$ Machine learning (98)
	$C_{4,3}$ Deep learning (9)	$C_{4,4}$ Natural language processing (11)
	$C_{4,5}$ Others (4)	
D_5 Delivery channel	$C_{5,1}$ (Cloud) platform (165)	$C_{5,2}$ Software-as-a-service (43)
	$C_{5,3}$ Infrastructure-as-a-service (3)	$C_{5,4}$ Others (18)
D_6 Detection environment	$C_{6,1}$ Business internal systems (106)	$C_{6,2}$ Internet (9)
	$C_{6,3}$ Business internal systems+internet (84)	$C_{6,4}$ Social media+dark and deep web (10)
	$C_{6,5}$ Physical things (4)	$C_{6,6}$ Multiple (14)
	$C_{6,7}$ Others (2)	
D_7 Cyber risk	$C_{7,1}$ Third-party (160)	$C_{7,2}$ First party (7)
	$C_{7,3}$ Both (62)	
D_8 Response strategy	$C_{8,1}$ No (26)	$C_{8,2}$ Automated (87)
	$C_{8,3}$ Manual (116)	
D_9 Concept of security information	$C_{9,1}$ Threat vulnerability (105)	$C_{9,2}$ Attack (35)
	$C_{9,3}$ Threat vulnerability+impact (15)	$C_{9,4}$ Attack+impact (8)
	$C_{9,5}$ Threat vulnerability+ attack+impact (26)	$C_{9,6}$ Threat vulnerability+impact+controls (6)
	$C_{9,7}$ Attack+impact+controls (5)	$C_{9,8}$ Multiple (29)
D_{10} CIA triad	$C_{10,1}$ Confidentiality (63)	$C_{10,2}$ Integrity (4)
	$C_{10,3}$ Availability (38)	$C_{10,4}$ Multiple (124)
D_{11} Alert of incident	$C_{11,1}$ Yes (138)	$C_{11,2}$ No (91)
D_{12} Priorization of threats	$C_{12,1}$ Yes (89)	$C_{12,2}$ No (140)
D_{13} Consulting service	$C_{13,1}$ Yes (58)	$C_{13,2}$ No (171)

Table 2. Final Taxonomy (number of services in brackets)

AI-driven Cybersecurity Business Models Archetypes

Table 3 shows the cluster analysis results and highlights the percentage distribution of the characteristics in the four archetypes. Each characteristic is labeled in color, with 0% in white and 100% in dark gray. For example, the *protection target* in Archetype 3 consists of 80% the *cyberspace*. Each cluster is listed in a column and can be interpreted as an archetype with different attributes. In addition, the percentage distribution of all classified services is shown between the dimensions and characteristics column. The underlying dataset can be found in the online Appendix.

Archetype 1 - Intrusion Detection and Resilience-Enhancing Cybersecurity Services. In the largest archetype, which consists of 80 services, intrusion detection and resilience-enhancing cybersecurity solutions are offered that provide less response to incidents. The key objective of these services is to identify vulnerabilities and threats. This involves scanning business internal systems and the internet.

Dimension	Σ	Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	n=229		n=80	n=38	n=50	n=61
D_1 Target group	3%	$C_{1,1}$ Individuals	4%	3%	6%	2%
	80%	$C_{1,2}$ Business	90%	55%	66%	95%
	1%	$C_{1,3}$ Governments	3%	0%	0%	0%
	11%	$C_{1,4}$ Business+governments	3%	32%	22%	2%
	4%	$C_{1,5}$ Multiple	1%	11%	6%	2%
D_2 Protection target	53%	$C_{2,1}$ Cyberspace	58%	42%	80%	31%
	2%	$C_{2,2}$ Physical space	5%	0%	2%	0%
	29%	$C_{2,3}$ Cyber+physical space	28%	29%	6%	51%
	1%	$C_{2,4}$ Brand protection	0%	0%	2%	2%
	14%	$C_{2,5}$ Multiple	10%	29%	10%	15%
	1%	$C_{2,6}$ Others	0%	0%	0%	2%
D_3 Service provided	24%	$C_{3,1}$ Resilience	40%	0%	38%	7%
	3%	$C_{3,2}$ Cyberattack prevention	1%	0%	10%	0%
	10%	$C_{3,3}$ Intrusion detection	24%	0%	6%	0%
	14%	$C_{3,4}$ Cyberattack prevention+intrusion detection	16%	18%	18%	3%
	4%	$C_{3,5}$ Cyberattack prevention+response	3%	5%	8%	3%
	8%	$C_{3,6}$ Intrusion detection+response	10%	0%	12%	8%
	39%	$C_{3,7}$ Multiple	8%	76%	10%	82%
D_4 AI technology	46%	$C_{4,1}$ No specified AI	60%	8%	62%	39%
	43%	$C_{4,2}$ Machine learning	28%	84%	28%	49%
	4%	$C_{4,3}$ Deep learning	1%	8%	2%	7%
	5%	$C_{4,4}$ Natural language processing	5%	0%	8%	5%
	2%	$C_{4,5}$ Others	5%	0%	0%	0%
D_5 Delivery channel	72%	$C_{5,1}$ Platform	78%	26%	70%	95%
	19%	$C_{5,2}$ Software-as-a-service	8%	61%	26%	2%
	1%	$C_{5,3}$ Infrastructure-as-a-service	0%	8%	0%	0%
	8%	$C_{5,4}$ Others	15%	5%	4%	3%
D_6 Detection environment	46%	$C_{6,1}$ Business internal systems	63%	18%	72%	21%
	4%	$C_{6,2}$ Internet	9%	0%	4%	0%
	37%	$C_{6,3}$ Business internal systems+internet	16%	74%	18%	56%
	4%	$C_{6,4}$ Social media+dark web+deep web	4%	3%	2%	8%
	2%	$C_{6,5}$ Physical devices	4%	0%	0%	2%
	6%	$C_{6,6}$ Multiple	5%	3%	2%	13%
	1%	$C_{6,7}$ Others	0%	3%	2%	0%
D_7 Cyber risk	70%	$C_{7,1}$ Third party	81%	24%	60%	92%
	3%	$C_{7,2}$ First party	5%	0%	6%	0%
	27%	$C_{7,3}$ Both	13%	76%	34%	8%
D_8 Response strategy	38%	$C_{8,1}$ No	79%	3%	46%	0%
	50%	$C_{8,2}$ Automated	16%	74%	30%	98%
	12%	$C_{8,3}$ Manual	5%	24%	24%	2%
D_9 Concept of security information	46%	$C_{9,1}$ Threat vulnerability	81%	0%	58%	18%
	15%	$C_{9,2}$ Attack	10%	3%	6%	38%
	7%	$C_{9,3}$ Threat vulnerability+impact	4%	0%	14%	8%
	3%	$C_{9,4}$ Attack+impact	0%	11%	6%	2%
	11%	$C_{9,5}$ Threat vulnerability+attack+impact	4%	26%	2%	20%
	2%	$C_{9,6}$ Threat vulnerability+impact+controls	0%	8%	2%	2%
	3%	$C_{9,7}$ Attack+impact+controls	0%	5%	6%	2%
	13%	$C_{9,8}$ Multiple	1%	47%	6%	11%
D_{10} CIA triad	28%	$C_{10,1}$ Confidentiality	34%	13%	60%	2%
	2%	$C_{10,2}$ Integrity	3%	3%	2%	0%
	17%	$C_{10,3}$ Availability	41%	0%	10%	0%
	54%	$C_{10,4}$ Multiple	24%	84%	26%	98%
D_{11} Alert of incident	60%	$C_{11,1}$ Yes	0%	100%	96%	85%
	40%	$C_{11,2}$ No	100%	0%	4%	15%
D_{12} Prioritization of threats	39%	$C_{12,1}$ Yes	2%	97%	52%	39%
	61%	$C_{12,2}$ No	98%	3%	48%	59%
D_{13} Consulting services	25%	$C_{13,1}$ Yes	14%	58%	22%	23%
	75%	$C_{13,2}$ No	86%	42%	78%	77%

Table 3. Cluster Analysis Results

Intrusion detection and resilience-enhancing cybersecurity services include, for example, (Dathena 2022; OneLogin 2022; Spin ai 2022). Dathena utilizes AI-driven engines to identify and classify the risk exposure of sensitive data. This can make data exchange management more secure. Spin.ai provides ransomware monitoring, risk assessments, and backup and recovery solutions for Google G Suite and Microsoft Office 365 environments on a platform called 'SpinOne'. OneLogin offers a platform to manage digital identities for employees and customers by using contextual authentication requirements. They identify and analyze potential threats by utilizing proprietary ML. Services in Archetype 1 do not provide an alert if an incident occurs. Also, there is no prioritization of threats, and no consulting service.

Archetype 2 - All-In-One-Solutions. The smallest Archetype (n=38 services) provides multiple offerings, including resilience, cyberattack prevention, intrusion detection, and response in one service. ML technologies are utilized to detect and respond to anomalies and attacks in real-time (e.g., GroupSense and 24metrics). However, it also sets up cyber traps by creating realistic simulations of data and systems. The company Penten offers this cyber deception service designed to defend against and deceive highly advanced threats (Penten 2022). It protects cyber risk against first- and third-party attacks and addresses the multiple objectives of information security (CIA). All services send out an alert when an incident occurs and in 97% of cases have either a manual or automatic response strategy. In addition, the threats are prioritized. 24metrics provides a software-as-a-service (SaaS) solution to prevent and detect threats. This scans conversions for bots, click spam, duplicate internet protocol, fraud, and duplicate logins, and prevents them with real-time blocking (24metrics 2022). Groupsense offers digital risk protection in the form of threat monitoring and mitigation, risk mapping, and prioritization across several environments such as dark and deep web and business internal systems (GroupSense 2022).

Archetype 3 - Governance and Compliance Enhancing Services. Business models in Archetype 3 (n= 50 services) provide measures to ensure and support governance and compliance. For this, cyberspace is protected to secure confidentiality. The detection environment extends across business internal systems and the internet. Typical functions these services offer are data management (e.g., Kriptos), risk analysis, penetration testing (e.g., SEWORKS), logins, and two-factor authentication. Kriptos provides a service that automatically classifies unstructured data by using AI. Sensitive documents, personal data, confidential information, and credit cards can be identified, and access to those data is restricted to authorized users in real-time (Kriptos 2022). SEWORKS provides an automated penetration testing solution by simulating real-world attack scenarios to identify cybersecurity vulnerabilities (SEWORKS 2022).

Archetype 4 - Third-Party Attack Prevention with Automated Response. The second-largest Archetype, with 61 services, offers attack prevention solutions with an automated response. The detection environment spans business internal systems, internet, social media, and dark and deep web. The services protect threat of third-party attacks and addresses the multiple objectives of information security (CIA). The most common attacks protected by this business model are phishing attacks (e.g., PhishFirewall), ransomware and data leaks (e.g., Cybelangel), bot attacks, and disinformation. PhishFirewall offers cybersecurity awareness education by emulating threats and analyzing the human security factor (PhishFirewall 2022). Cybelangel provides data leak and ransomware detection by scanning the surface, dark and deep web, connected storage devices, open databases, and cloud apps (Cybelangel 2022).

Towards a Decision Tree for AI-Driven Cybersecurity (DETRAICS)

Based on the taxonomy and the identified archetypes, we build the decision tree *DETRAICS*. It is possible to obtain a specific archetype recommendation by answering four questions (see Figure 2). DETRAICS uses the dimensions from the taxonomy (see Table 2) as questions and their corresponding characteristics as answers. The first question asks if the service should employ a response strategy to a cyber incident. This question can be answered with "Yes" or "No." If a response strategy should be employed, then the left path of the tree needs to be followed. If no response strategy should be employed, then the right path of the tree needs to be followed. On both sides of the tree, the question of cyber risk is presented next. This question can be answered either with third-party risk or first- and third-party risk. For better understanding, we explain the leftmost and rightmost paths of DETRAICS. If the first answer indicates that the service should employ a response strategy and the cyber risk represents a third-party risk, then the question is which security information goal should be addressed. This question can be answered with either confidentiality or availability, or multiple goals. If the question is answered with multiple goals, then the next question asks what type of response strategy should be executed. This can either be automatic or manual. If the response strategy should be employed automatically, then Archetype 4 is recommended, if the response strategy should be employed manual, then Archetype 1 is recommended. If the first answer indicated that the service should not employ a response strategy and the cyber risk represents a third-party risk, the next question asks which target group is addressed. This can either be answered with individual or business and government or solely business. If the target group is business, the next question asks which concept of security information should be applied. This can be answered with the representation of attacks and the impact or the representation of threat vulnerabilities. If the latter should be represented, then Archetype 1 is recommended and if attacks and their impact should be represented, then Archetype 3 is recommended.

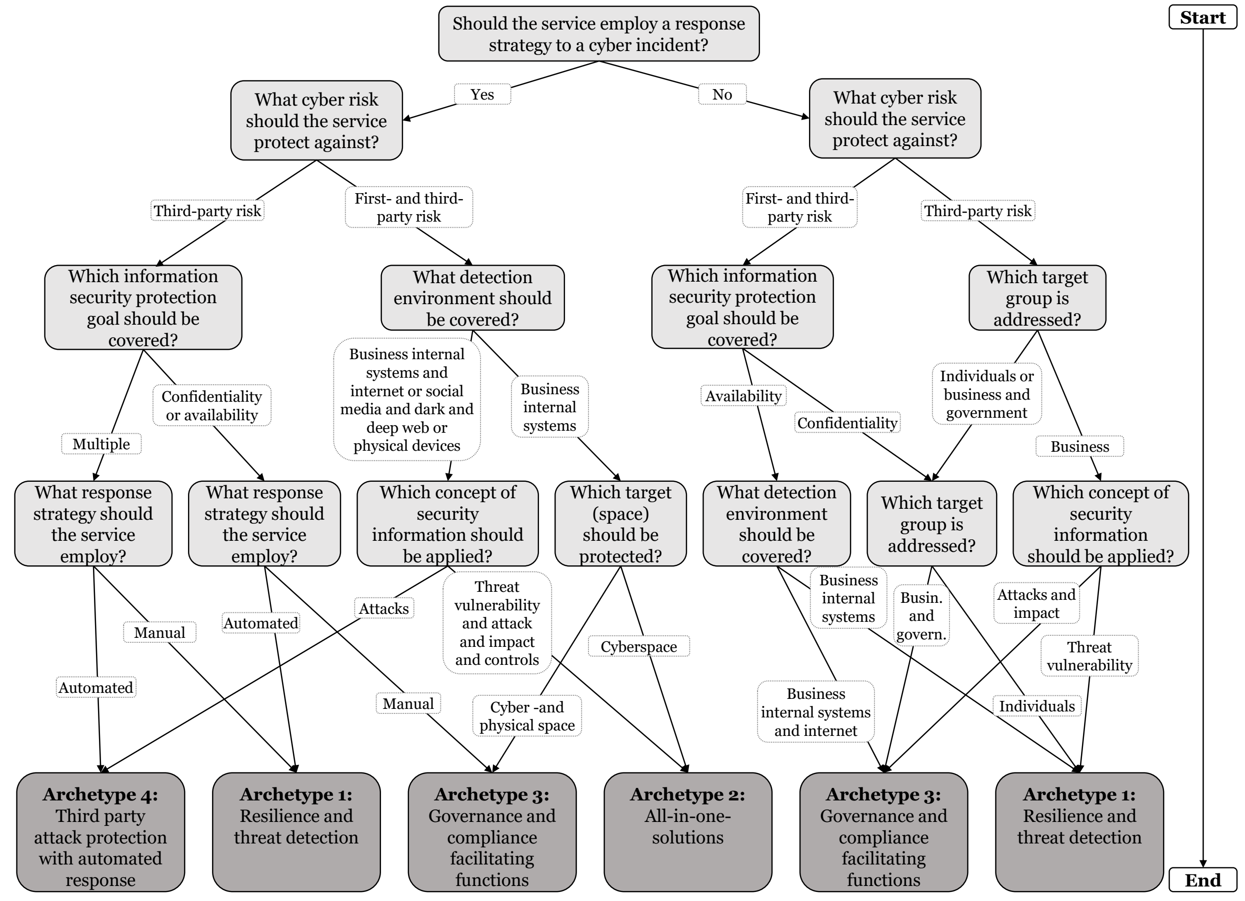
Start
Should the service employ a response strategy to a cyber incident?
Yes
No
What cyber risk should the service protect against?
What cyber risk should the service protect against?
Third-party risk
First- and third-party risk
First- and third-party risk
Third-party risk
Which information security protection goal should be covered?
What detection environment should be covered?
Which information security protection goal should be covered?
Which target group is addressed?
Confidentiality or availability
Multiple
Business internal systems and internet or social media and dark and deep web or physical devices
Business internal systems
Availability
Confidentiality
Individuals or business and government
Business
What response strategy should the service employ?
What response strategy should the service employ?
Which concept of security information should be applied?
Which target (space) should be protected?
What detection environment should be covered?
Which target group is addressed?
Which concept of security information should be applied?
Manual
Automated
Automated
Attacks
Threat vulnerability and attack and impact and controls
Manual
Cyber -and physical space
Cyberspace
Business internal systems
Busin. and govern.
Business internal systems and internet
Attacks and impact
Threat vulnerability
Individuals
Archetype 4: Third party attack protection with automated response
Archetype 1: Resilience and threat detection
Archetype 3: Governance and compliance facilitating functions
Archetype 2: All-in-one-solutions
Archetype 3: Governance and compliance facilitating functions
Archetype 1: Resilience and threat detection
End

Figure 2. Decision Tree for AI-driven Cybersecurity (DETRAICS)

Discussion and Evaluation

We developed a taxonomy based on scientific literature and empirical data of 229 real-world services which allows us to classify AI-driven cybersecurity services. Based on this, we created DETRAICS. All three results together contribute to theory and practice, but also provide a contribution on its own. While other scientists have focused on the theoretical (Sarker et al. 2021) and technical (Xin et al. 2018) implementation of AI-driven cybersecurity, we have focused on the empirically-based considerations and contributions to the underlying business models and services provided. Regarding the 13 examined dimensions and 58 characteristics, we identified that there are large differences in the distribution frequency of characteristics within dimensions, see Table 2 and 3. For example, 50% of all services offer an automated response strategy, 12% a manual, and 38% offer no response strategy. These distributions allow practitioners to identify design elements and make decisions for cybersecurity implementation in the corporate ecosystem. However, it also allows researchers to identify possible technology trends and adoptions. Our key findings indicate that considering the distribution frequency of the characteristics of all services (Table 3), an average AI-driven cybersecurity solution based on the majority of services addresses businesses to protect the cyberspace against external (third-party) threats by applying ML techniques that scans business internal systems and the internet, provided on a platform. The services differ in the classification of dimensions and characteristics, which allows us to deduce four archetypes, leading to specific business models with different goals. When the goal of the service is to enhance the resilience and detect threats, services in Archetype 1 have to be selected. If multiple offerings, including resilience, cyberattack prevention, intrusion detection, and response, are required, Archetype 2 must be selected. To facilitate governance functions, such as penetration tests or multi-factor authentication, services in Archetype 3 have to be selected. When third-party risk, especially ransomware and phishing, is targeted with automated response processes, services in Archetype 4 must be selected.

We added dimensions and characteristics during the taxonomy development process on later iterations. This includes the dimension *D_6 Detection environment*. We first added this dimension in the fourth E2C-iteration as we identified that many services clearly illustrated which environment scanning and monitoring processes are conducted to search for malicious activities. Besides the detection environment of business internal systems by default, we identified other characteristics, including social media, dark web, and deep web. Such environments are particularly examined because insider threats, character assassination, and misinformation can be initialized by malicious (corporate internal) actors on those non-free accessible platforms. *$C_{2,4}$ Brand protection* in the dimension *D_2 Protection target* is a further characteristic we added during the later taxonomy development iterations. The brand protection service of Quointelligence monitors the threat exposure of the brand by identifying trademark abuse, data leaks, and domains masquerading as a brand on the internet. Such threats can be originated from an exposed asset or hacked password (Quointelligence 2022). However, the brand protection characteristic is mostly interconnected with the *Cyberspace* ($C_{2,1}$) characteristic, resulting in multiple characteristics ($C_{2,5}$). Kumar et al. (2020) proposed different application areas for cybersecurity solutions. For this reason, in the initial dimension in the first C2E iteration, we set up two different dimensions, one for the sectors and application areas in cyberspace and one for the physical space. According to Kumar et al. (2020), in cyberspace, or cyber-tier, various areas need to be protected, such as data centers, servers, mobile devices, workstations, and cloud storage. The physical space includes the areas of critical infrastructure in particular, such as manufacturing, power grids, nuclear reactors, healthcare, and transportation. In further iterations, we have identified no sector delineation in real-world services. The services were merely cross-sectoral and thus did not address any particular target group as a sector. Here, we identified that the worlds between academic literature and real-world applications blur. Determining the *Target group (D_1)* and the type of service used also proved to be a challenge, which is why we only identified three groups: individuals, businesses, and governments.

Considering the AI-driven component of the cybersecurity services, applying this technology can have positive and negative impacts according to self-learning, automation, efficiency, explainability, data ethics, and privacy aspects. Regarding the positive impacts, ML methods can be applied to expose and block attacks, conduct threat analysis and forensics, and stop ransomware and zero-day threats (Deep Instinct 2022; Traceble AI 2022). The analyzed services use multiple ML techniques such as deep learning, clustering, classification, and unsupervised learning (Traceble AI 2022). The underlying model has the ability to learn and improve over time to increase the efficiency of cyberattack prevention. So, the models are self-learning, autonomously predicting, detecting, and preventing threats without the necessity to update for maintenance (Deep Instinct 2022). However, based on the taxonomy and the archetype analysis, a large variability of AI-driven cybersecurity services can be identified. Nevertheless, responsible stakeholders need to determine whether AI technologies must be used for all cybersecurity needs. The usefulness and efficiency of the technologies and their contribution to cybersecurity must be carefully evaluated individually at first hand. The more data is processed and trained in the model, the more efficient the results are. But, this growing need to train data raises concerns regarding ethics, data protection, and privacy (Berente et al. 2021). The more data that is processed, the more patterns can be identified from this data. It is uncertain whether these patterns can also be misused and whether the intended increase in cybersecurity can lead in the opposite direction. According to Berente et al. (2021) the capability of AI emulations allows to distinguish bot behavior and human behavior, characterized by human errors and biases. As a consequence, hackers can exploit such AI-driven detection by including human behavior into malicious code on purpose. This risk must be weighed against the benefits that this service provides.

Several studies established guidelines for ethical AI to address concerns about the negative side effects of using AI-driven solutions, including privacy and human rights violations, wrong and biased decision-making due to incomprehensible non-transparent algorithms (Mayer et al. 2021; Seppälä et al. 2021).

Guidelines for ethical AI call for increased fairness, non-discrimination, responsibility, accountability, transparency, and explainability. To manage fairness and non-discrimination of AI systems, data governance can be practiced. This includes the concept of data minimization along the whole data life cycle by collecting, using, and storing as fewer data as necessary (Kroll 2018; Seppälä et al. 2021). Responsibility and accountability for the technology and its actions must be allocated to responsible stakeholders who develop, provide, and use the system. Fjeld et al. (2020) proposed that the implementation of impact assessments and the ability of humans to control and audit the systems must be guaranteed permanent. However, the human ability to audit such systems decreases as the services become more intertwined with the corporate system. Darktrace promotes their self-learning, autonomous, and real-time acting so-called 'Darktrace Immune System,' which can access all network devices, clouds, e-mails, and SaaS (Darktrace 2022). However, the degree of human accountability is questionable. Transparency and explainability can be implemented by minimizing black boxes and increasing the degree of interpretable results (Seppälä et al. 2021). In the analysis of the services, we determined that a high value was placed on the explainable visualization of the results. Further transparency can be achieved by presenting the impact of the threats and vulnerabilities using graphs and highlighting the threat situation in red, yellow, and green, for example.

Regarding the interviews, we found mostly positive statements for the taxonomy, the archetypes, and DETRAICS. Focusing on the completeness of the artifacts, Experts 5, 6, and 7 recommended including the dimensions pricing, and market size into the taxonomy. They pointed out that the market competition should be reflected. Moreover, for clients of a consulting company, it is important to know which (big) players are relevant and which are not. Due to the lack of information on pricing and market size for all 229 services, it was not possible to include such information in the taxonomy yet. Usefulness and appropriateness were broadly confirmed by the interviewees. Moreover, Experts 5, 6, and 7 suggested portfolio and investment managers as potential users of our artifacts. Some statements were given concerning the practical application of the taxonomy, the archetypes, and DETRAICS. Experts 1 and 2, who directly work in an IT department, can and will use our results and findings for strategic planning. This includes their company's cybersecurity strategy alignment in the mid- and long-term. Experts 3 and 4, who work for many different clients, found that our results are useful and beneficial for daily consulting purposes. They stated that our taxonomy could serve as a standard to design cybersecurity services, increase their professional knowledge, and raise clients' cybersecurity awareness. In addition, Experts 3 and 4 stated that DETRAICS could help in counseling activities within the scope of a client's cybersecurity check. Experts 5, 6, and 7 suggested that a search platform can use our taxonomy and give services the possibility to classify themselves. Based on this, a decision tree can be created, and investors and customers can receive advice on a suitable service.

Theoretical Contributions and Practical Implications

We provide three main contributions: First, we examined an empirically validated taxonomy of AI-driven cybersecurity business models. We identified 13 dimensions and 58 characteristics among 229 companies. Therefore, we provided a comprehensive and empirically based knowledge foundation for academics and followed the further research directions by Wallace et al. (2020) and Sarker et al. (2021). Second, we identified four specific archetypes among the investigated companies. Using clustering techniques, we identify additional information between the similarities and differences of the services provided, which a taxonomy in its single, descriptive form cannot achieve (Möller et al. 2021). By successfully examining archetypes, we evaluate the information identified with the taxonomy development, as proposed by Kundisch et al. (2021). Our theoretical contributions are both the taxonomy dimensions and characteristics and the specific archetypes. Taxonomies can be a starting point and meaningful knowledge foundation for theory-building purposes, like design theories (Muntermann et al. 2015; Kundisch et al. 2021). In addition, the taxonomy can be used as a glossary with important and domain specific vocabulary (Weking et al. 2020). Based on the identified business models we build on business model literature in the emerging field of AI-driven cybersecurity services, which is a white spot according to Möller et al. (2021). We strengthen the knowledge of AI-driven cybersecurity business models and its value creation by integrating practical insights and deducing archetypes. Our archetypes can be used for more tailored investigations of critical business model elements and their relationships. We encourage researchers to undertake more research in this area and provide further research directions. Our research additionally offers opportunities to develop a decision tree in other research areas, especially where strategic planning is required. Thus, we show how to extend taxonomy and archetype analysis. The interviews showed several possibilities on how the results of this study can be used by practitioners. Academics can use these observations and keep them in mind as a starting point, e.g., in a C2E-iteration for their own taxonomy, archetype, and decision tree development.

Cybersecurity companies, on the other hand, can use the findings provided. They can see what the market is doing and modify their offerings. We ignite a broader discussion among academics and practitioners on the crucial business model components of AI-driven cybersecurity solutions. Third, we propose *DETRAICS*, a decision tree for AI-driven cybersecurity services, as a practical implication. Stakeholders can use *DETRAICS* to investigate alternatives for implementing AI-driven cybersecurity solutions. Also, it can be used as a quick hands-on decision-support tool by responsible people. While the market for AI is rising and can be overwhelming for stakeholders, our decision tree reduces the complexity of the market towards a meaningful, easy-to-use decision tool. Furthermore, it is visualized intuitively and lowers the weaknesses that taxonomies and archetypes can be seen as too academic for practitioners. However, the archetypes can have advantages and disadvantages considering the necessity and usefulness of individual requirements, for example, corporate needs, critical infrastructure stakeholders, efficiency, monetary, or insurance purposes. *DETRAICS* also shows which of the 13 dimensions are crucial factors defining the archetype and which dimensions are closely interrelated. Looking at *DETRAICS,* the most influencing dimensions are the *Response strategy*, the *Cyber risk*, the *CIA triad*, and the *Detection environment*, as these questions have been asked frequently and on the top of the decision tree. This led to a reduction of complexity.

Limitations, Further Research, and Conclusions

We evaluated the taxonomy, the archetypes, and DETRAICS with a first set of expert interviews. Since taxonomies and their outputs are classified as Design Science Research artifacts, their problem-solving nature can be further examined. Therefore, constant observation of the usefulness with, e.g., (semi-structured) interviews with intended target groups are advisable for further evaluation. This leads to additional development iterations and can further improve our problem-solving artifacts. Data collection of our taxonomy building procedure through Crunchbase took place in March 2022. As indicated in the theoretical background section, the market for AI-driven cybersecurity solutions will rise. We present a market snapshot, but a continuous market observation is necessary. Further research can add new objects to the taxonomy or delete or reframe existing ones. Since taxonomies are extendable from their nature, advancements are always possible (Nickerson et al. 2013). New objects, or the deletion of objects because of, e.g., a merger or bankruptcy, can lead to new archetypes and a new decision tree. Researchers can further investigate the relationships between specific constructs, i.e., dimensions and characteristics of the business models and underlying services. Different combinations of business model components can be further investigated, for example, with the Five-V framework by Taran et al. (2016). This framework assists business model researchers in finding research contributions regarding (successful) business model configurations and can advance the academic body of knowledge. Deducing four specific business models, each one can be examined in more detail and, for example, respective taxonomies can be created. Further research can build on our taxonomy, business model archetypes, and DETRAICS to develop a maturity model or to deduce design principles for cybersecurity solution development. In addition, while we have not found information about customer satisfaction with the services, case studies and data mining analyses with subsequent sentiment assessments can be performed. In a further case study, it can be investigated whether and in which areas differences between AI-driven and non-AI-driven cybersecurity services are detectable and where value is created. We also offer potential for application in a variety of areas.

We developed our taxonomy of AI-driven cybersecurity business models and services. With taxonomy development methods proposed by Nickerson et al. (2013) and Kundisch et al. (2021), we identify 13 dimensions and 58 characteristics of analyzed 229 services. With this taxonomy, we answered RQ1 presenting four specific archetypes that serve as a meaningful foundation for a discussion among academics and practitioners. By answering RQ2, we propose *DETRAICS*, a decision tree for AI-driven cybersecurity services, based on the identified dimensions and characteristics, and archetypes. *DETRAICS* can be used by decision-makers to investigate alternatives and to choose adequate cybersecurity services.

Acknowledgements

This research was funded by the Lower Saxony Ministry of Science and Culture under grant number 11-76251-13-3/19–ZN3488 (ZLE) as well as 11-76251-15-1/19–ZN3563 (SiNED) within the Lower Saxony

"Vorab" of the Volkswagen Foundation. It was supported by the Center for Digital Innovations (ZDIN) and the Energy Research Centre of Lower Saxony (EFZN).

References

24metrics 2022. "*Conversion Screening & A.I Quality Predictions*," Retrieved online on September 5, 2022 from https://www.24metrics.com/.

Aftergood, S. 2017. "Cybersecurity: The Cold War Online," *Nature* (7661), pp. 30-31.

Albus, J. S. 1991. "Outline for a Theory of Intelligence," *IEEE Transactions on Systems, Man, and Cybernetics* (21:3), pp. 473–509.

Alsayed, A., and Bilgrami, A. 2017. "E-Banking Security: Internet Hacking, Phishing Attacks, Analysis and Prevention of Fraudulent Activities," *International Journal of Emerging Technology and Advanced Engineering* (7:1), pp. 109–115.

Babatunde, D. E., Anozie, A. N., and Omoleye, J. 2020. "Artificial Neural Network and its Applications in the Energy Sector–An Overview," *International Journal of Energy Economics and Policy* (10:2), pp. 250–264.

Berente, N., Gu, B., Recker, J., and Santhanam, R. 2021. "Managing Artificial Intelligence," *MIS Quarterly* (45:3), pp. 1433-1450.

CISA 2022. "*Cybersecurity*," Retrieved online on September 5, 2022 from https://www.cisa.gov/cybersecurity.

Cowbell Cyber 2022. "*About Cowbell*," Retrieved online on September 5, 2022 from https://cowbell.insure/about-cowbell/.

Craigen D, Diakun-Thibault N, and Purse R. 2014. "Defining Cybersecurity," *Technology Innovation Management Review* (4:10), pp. 13–21.

Croasdell, D., and Palustre, A. 2019. "Transnational cooperation in cybersecurity," in *Proceedings of the 52nd Hawaii International Conference on System Sciences.*

Crunchbase 2022. "*Crunchbase*," Retrieved online on September 5, 2022 from https://www.crunchbase.com/.

Cybelangel 2022. "*About us*," Retrieved online on September 5, 2022 from https://cybelangel.com/about-us/.

Enisa 2022. "*Big Data*," Retrieved online on September 5, 2022 from https://www.enisa.europa.eu/topics/cloud-and-big-data/big-data.

European Parliament 2022a. "*Digital Transformation: Importance, Benefits and EU Policy*," Retrieved online on September 5, 2022 from https://www.europarl.europa.eu/news/en/headlines/priorities/digital-transformation/20210414STO02010/digital-transformation-importance-benefits-and-eu-policy.

European Parliament 2022b. "*Cybersecurity: Main and Emerging Threats in 2021*," Retrieved online on September 5, 2022 from https://www.europarl.europa.eu/news/en/headlines/society/20220120STO21428/cybersecurity-main-and-emerging-threats-in-2021-infographic.

Darktrace 2022. "*About the Company*," Retrieved online on September 5, 2022 from https://www.darktrace.com/en/overview/.

Dathena 2022. "*About us*," Retrieved online on September 5, 2022 from https://www.dathena.io/company/about.

Deep Instinct 2022. "*Why deep instinct*," Retrieved online on September 5, 2022 from https://www.deepinstinct.com/why-deep-instinct.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, Á., and Srikumar, M. 2020. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," *SSRN Electronic Journal.*

GroupSense 2022. "*Company*," Retrieved online on September 5, 2022 from https://www.groupsense.io/company.

Hansen, J. V., Lowry, P. B., Meservy, R. D., and McDonald, D. M. 2007. "Genetic Programming for Prevention of Cyberterrorism through Dynamic and Evolving Intrusion Detection," *Decision Support Systems* (43:4), pp. 1362–1374.

Jang-Jaccard, J., and Nepal, S. 2014. "A Survey of Emerging Threats in Cybersecurity," *Journal of Computer and System Sciences (*80:5), pp. 973–993.

Jöhnk, J., Weißert, M., and Wyrtki, K. 2021. "Ready or Not, AI Comes-An Interview Study of Organizational AI Readiness Factors," *Business & Information Systems Engineering* (63:1), pp. 5–20.

Kaufman, L., and Rousseeuw, P. J. 1990. *Finding Groups in Data*, Hoboken, NJ, USA: Wiley & Sons.

Kriptos 2022. "*Home*," Retrieved online on September 5, 2022 from https://www.kriptos.io/en/home.

Kroll, J. A. 2018. "Data Science Data Governance [AI Ethics]," *IEEE Security & Privacy* (16:6), pp. 61–70.

Kumar, S. L. 2017. "State of the Art-Intense Review on Artificial Intelligence Systems Application in Process Planning and Manufacturing," *Engineering Applications of Artificial Intelligence* (65), pp. 294–329.

Kumar, C., Marston, S., and Sen, R. 2020. "Cyber-Physical Systems (CPS) Security: State of the Art and Research Opportunities for Information Systems Academics," *Communications of the Association for Information Systems* (47), pp. 678–696.

Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., and Szopinski, D. 2021. "An Update for Taxonomy Designers," *Business & Information Systems Engineering* (online first), pp. 1–19.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep Learning," *Nature* (521), pp. 436–444.

Likas, A., Vlassis, N., and Verbeek, J. J. 2003. "The global k-means clustering algorithm," *Pattern recognition* (36:2), pp. 451-461.

Magee, J. F. 1964. "Decision Trees for Decision Making," *Harvard Business Review* (42), pp. 126–138.

Mayer, A.-S., Haimerl, A., Strich, F., and Fiedler, M. 2021. "How corporations encourage the implementation of AI ethics," in *Proceedings of the Twenty-Ninth European Conference on Information Systems*.

McIntosh, T., Jang-Jaccard, J., Watters, P., and Susnjak, T. 2019. "The Inadequacy of Entropy-Based Ransomware Detection," in *Proceedings of the International Conference on Neural Information Processing*.

Möller, F., Stachon, M., Azkan, C., Schoormann, T., and Otto, B. 2021. "Designing Business Model Taxonomies–Synthesis and Guidance from Information Systems Research," *Electronic Markets* (online first), pp. 1-26.

Moon, D., Im, H., Kim, I., and Park, J. H. 2017. "DTB-IDS: An Intrusion Detection System based on Decision Tree using Behavior Analysis for Preventing APT attacks," *The Journal of Supercomputing* (73:7), pp. 2881–2895.

Mueller, N. S., Werth, O., Koenig, C. M., and Breitner, M. H. 2022. "How is Your Mood Today? - A Taxonomy-based Analysis of Apps for Depression," in *Proceedings of the Twenty-eight Americas Conference on Information Systems*.

Mullet, V., Sondi, P., and Ramat, E. 2021. "A Review of Cybersecurity Guidelines for Manufacturing Factories in Industry 4.0," *IEEE Access* (9), pp. 23235–23263.

Muntermann, J., Nickerson, R., and Varshney, U. 2015. "Towards the Development of a Taxonomic Theory," in *Proceedings of the Twenty-first Americas Conference on Information Systems*.

Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A Method for Taxonomy Development and Its Application in Information Systems," *European Journal of Information Systems* (22:3), pp. 336–359.

Nweke, L. O. 2017. "Using the CIA and AAA Models to Explain Cybersecurity Activities," *PM World Journal* (6:12), pp. 1–2.

Ohm, M., Plate, H., Sykosch, A., and Meier, M. 2020. "Backstabber's Knife Collection: A Review of Open Source Software Supply Chain Attacks," in *Proceedings of the International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*.

OneLogin 2022. "*Why onelogin*," Retrieved online on September 5, 2022 from https://www.onelogin.com/why-onelogin.

Padmanabhan, B., Sahoo, N., and Burton-Jones, A. 2022. "Machine Learning in Information Systems Research," *MIS Quarterly* (46:1), pp. iii–xix.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., Weiss, R.J., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. 2011. "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* (12), pp. 2825–2830.

Penten 2022. "*About penten*," Retrieved online on September 5, 2022 from https://www.penten.com/about/.

Phishfirewall 2022. "*About us,*" Retrieved online on September 5, 2022 from https://www.phishfirewall.com/about-us/.

Pillsbury 2021. "*AI & Cybersecurity: Balancing Innovation, Execution & Risk*," Retrieved online on September 5, 2022 from https://www.pillsburylaw.com/en/news-and-insights/ai-and-cybersecurity-balancing-innovation-execution-and-risk.html.

Punj, G., and Stewart, D. W. 1983. "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research* (20:2), pp. 134–148.

Quointelligence 2022. "*Tailor-Made Threat Intelligence*," Retrieved online on September 5, 2022 from https://quointelligence.eu/.

Saputra, D. M., Saputra, D., and Oswari, L. D. 2020. "Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method," in *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications*.

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., and Ng, A. 2020. "Cybersecurity Data Science: An Overview from Machine Learning Perspective," *Journal of Big Data* (7:1), pp. 1–29.

Sarker, I. H., Furhad, M. H., and Nowrozy, R. 2021. "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions," *SN Computer Science* (2), pp. 1–18.

Seppälä, A., Birkstedt, T., and Mäntymäki, M. 2021. "From Ethical AI Principles to Governed AI," in *Proceedings of the Forty-Second International Conference on Information Systems*.

Schöbel, S. M., Janson, A., and Söllner, M. 2020. "Capturing the Complexity of Gamification Elements: A Holistic Approach for Analysing Existing and Deriving Novel Gamification Designs," *European Journal of Information Systems* (29:6), pp. 641–668.

SEWORKS 2022. "*About*," Retrieved online on September 5, 2022 from https://se.works/.

Sharifi, A. M., Amirgholipour, S. K., and Pourebrahimi, A. 2015. "Intrusion Detection Based on Joint of K-means and Knn," *Journal of Convergence Information Technology* (10:5), pp. 42–52.

Shaw, A. 2010. "Data Breach: From Notification to Prevention using pcidss," *Columbia Journal of Law and Social Problems* (43:4), pp. 517–562.

Spin ai 2022. "*About us*," Retrieved online on September 5, 2022 from https://spin.ai/company/about-us/.

Statista 2022. "*Artificial Intelligence (AI) in Cyber Security Market Value Worldwide from 2019 to 2027*," Retrieved online on September 5, 2022 from https://www.statista.com/statistics/1291380/ai-in-cyber-security-market-size/.

Sun, N., Zhang J., Rimba P., Gao S., Zhang L. Y., and Xiang Y. 2018. "Datadriven Cybersecurity Incident Prediction: A Survey," *IEEE Communications Surveys & Tutorials* (21:2), pp. 1744–1772.

Szopinski, D., Schoormann T., and Kundisch D. 2019. "Because Your Taxonomy is Worth it: Towards a Framework for Taxonomy Evaluation," in *Proceedings of the Twenty-Seventh European Conference on Information Systems*.

Taran, Y., Nielsen, C., Montemari, M., Thomsen, P., and Paolone, F. 2016. "Business Model Configurations: A Five-V Framework to Map Out Potential Innovation Routes," *European Journal of Innovation Management* (19:4), pp. 492–527.

Traceable AI 2022. "*How traceable AI works*," Retrieved online on September 5, 2022 from https://www.traceable.ai/how-it-works.

Wallace, S., Green, K., Johnson, C., Cooper, J., and Gilstrap, C. 2020. "An Extended TOE Framework for Cybersecurity Adoption Decisions," *Communications of the Association for Information Systems* (47:2020), pp. 338 – 363.

Warkentin, M., and Willison, R. 2009. "Behavioral and Policy Issues in Information Systems Security: The Insider Threat," *European Journal of Information Systems* (18:2), pp. 101–105.

Weber, M., Beutter, M., Weking, J., Böhm, M., and Krcmar, H. 2022. "AI Startup Business Models," *Business & Information Systems Engineering* (64:1), pp. 91–109.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. xiii–xxiii.

Weking, J., Stöcker, M., Kowalkiewicz, M., Böhm, M., and Krcmar, H. 2020. "Leveraging Industry 4.0–A Business Model Pattern Framework," *International Journal of Production Economics* (225), pp. 1–17.

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., and Wang, C. 2018. "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEE Access* (6), pp. 35365–35381.

Yin, C., Zhu, Y., Fei, J., and He, X. 2017. "A Deep Learning Approach for Intrusion Detection using Recurrent Neural Networks," *IEEE Access* (5), pp. 21954–21961.