Data Mining Methodology in Support of a Systematic Review of Human Aspects of Cybersecurity

Brendan M. Duffy[1,2(✉)] and Vincent G. Duffy[1(✉)]

[1] Purdue University, West Lafayette, IN 47907, USA
{duffy45,duffy}@purdue.edu
[2] West Lafayette High School, West Lafayette, USA

Abstract. Cybersecurity is an evolving field in the area of human-computer interactions (HCI), but human factors is a relevant area to consider when approaching cybersecurity. This report illustrates the findings of a systematic literature review of current publications on the emerging trends of human factors in cybersecurity. Analyses of content and bibliometrics were accomplished by using tools such as VOS Viewer, MAXQDA, Harzing, and AuthorMapper to establish the findings of emerging trends in the field. This report includes a step-by-step procedure for conducting the content analyses in each tool. The areas of human factors and cybersecurity are examined based on the data of the content analyses. A key finding is that human factors theory emerged from content analysis, and can be a basis for future research.

Keywords: Cybersecurity · Human factors · Bibliometric analysis

1 Introduction and Background

1.1 Cybersecurity and Human Factors

Cybersecurity is a field for protecting computers, all internet-capable systems, networks, servers, cloud data, and physical data from malicious software (malware) or cyber-attacks. The other field is Human Factors, which was concisely explained by The Human Factors and Ergonomics Society. The way they put it was, "Ergonomics and human factors use knowledge of human abilities and limitations to design systems, organizations, jobs, machines, tools, and consumer products for safe, efficient, and comfortable human use".

Cybersecurity began with the advent of internet development when users discovered flaws in system design. Cybersecurity was not being implemented until users with malicious intent began to take advantage of systems, due to a lack of protection against its usage for unintended purposes. In essence, the first notable case of a user exploiting system vulnerabilities and halt all internet processes with a worm.

The worm effectively incapacitated the functionality of the internet in 1988. The worm was created by Robert T. Morris, a Cornell student at the time. He created the

V. G. Duffy (Ed.): HCII 2020, LNCS 12199, pp. 242–253, 2020.
https://doi.org/10.1007/978-3-030-49907-5_17

worm to demonstrate the lack of security on computer networks. In turn, he was dismissed from Cornell and sentenced to three years of probation and was fined USD 10,050 (US. v R.T.Morris). Later on, he became a professor at MIT and was tenured in 2006 (MIT 2006). He was elected into the National Academy of Engineering in 2019 (NAE 2019). His work sparked an outbreak of worms and viruses, shortly after he had become notorious. These occurrences became the driving motivation for antivirus protection.

The heart of the issue was Human Factors because the companies did not consider all of the human processes of their systems. Their oversight caused human error and ill-intent to damage systems and other users. Human factors has been an emerging area in design since World War II-era because technological advances caused aircrafts to become more challenging to operate than the experienced pilots could manage. The military prompted engineers to design a more human-friendly system, which resulted in a lower fatality rate in combat, due to the improvement in design, based on human limitation.

1.2 Overview of Plan for Bibliometric Analysis and Systematic Review

A trend graph of papers involving both Human Factors and Cybersecurity was created with data Google Scholar. Metadata was retrieved from Google Scholar in Harzing. Then metadata was imported from the leading articles into VOS Viewer. Cluster analysis was completed using leading terms from each cluster. A co-citation analysis to find core articles was created with the reference data from the Web of Science in VOS Viewer. Content analysis was conducted on the core articles, based on the co-citation analysis. Leading articles from Google Scholar, ResearchGate, and Springerlink were used for a word cloud and lexical search in MAXQDA.

2 Purpose of Study

The objective of this study is to conduct a systematic literature review of papers on the topic of Human Factors in Cybersecurity. The review sought to summarize key aspects of new research in this emerging area and identify human factors that are forming the basis for further research on the topic of cybersecurity.

3 Research Methodology

3.1 Data Collection

To collect the data for the analyses, a keyword search was conducted in the two databases: Web of Science, and Google Scholar. There tends to be a higher volume of articles and papers in Google Scholar. The data from Web of Science includes title, abstract, authors, keywords, cited references, and source, although it has fewer articles and papers to analyze. Co-citation analyses require the reference data, which can only be extracted with computer assistance in the Web of Science database. The software, "Harzing's Publish or Perish" can extract bibliometric data from several databases, but is limited to 1000 articles from a search, and includes only the title, keywords, author, and source.

Because the search encompasses a vast summation of articles, it is the best method for collecting the essential keywords of the article sample. Google Scholar was used for the Harzing data extraction. The search terms that were used in the Web of Science and Harzing were, "cybersecurity AND "human factors" " and the search yielded 48 articles in the Web of Science, and was restricted to 1000 articles in Harzing (*Harzing's Publish or Perish*, n.d.). AuthorMapper has the Springer database of articles for the search terms.

3.2 Trend Analysis

The trend analysis is based on the results of the Web of Science data collection. Web of Science is equipped with a few tools for the analysis within the database, so those tools were used to analyze the trend data. All years were represented up to February 2020, and it shows the upward momentum of production in the literature on the topic.

Figure 1 shows the trend for articles involving both cybersecurity and human factors. The first listed publication was in 2014 in the Web of Science, and it shows a steady growth in the literature pool within the database from 2014 to 2020. The number of publications in the first two months of the year suggests that this year will have more publications than in 2019. It is especially important to consider the last 4 years when the number of publications per year stayed at a constant value, indicates that the field will grow from the plateau, or researchers will lose interest in the topic. The data from AuthorMapper gives a clearer picture (Fig. 2) AuthorMapper.

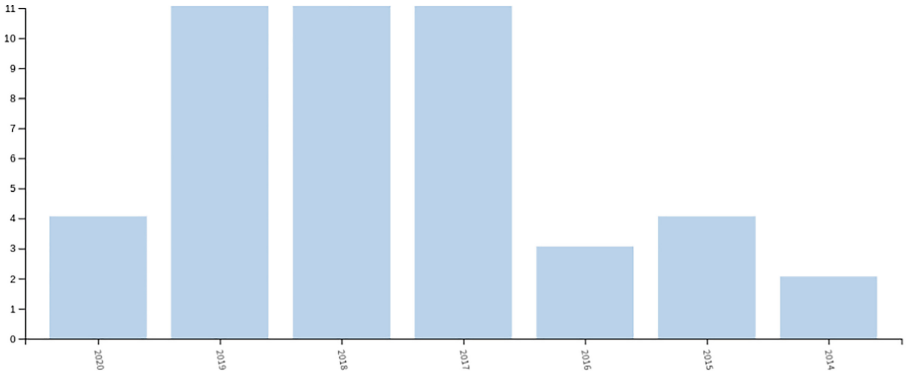

Fig. 1. Trend analysis of articles on cybersecurity and human factors (Web of Science, n.d.) The highest count on Y-axis is 11. X-axis starts with 2020 on the left and continues down to 2014.

It is clear, based on the trend analysis, that human factors in cybersecurity is a growing topic of research. The figure from AuthorMapper is very reassuring for the conclusion that the research is in an emerging area, especially after analyzing the graphs and seeing a jump in the volume of publications from 2018 to 2019, which was not represented in the Web of Science diagram.

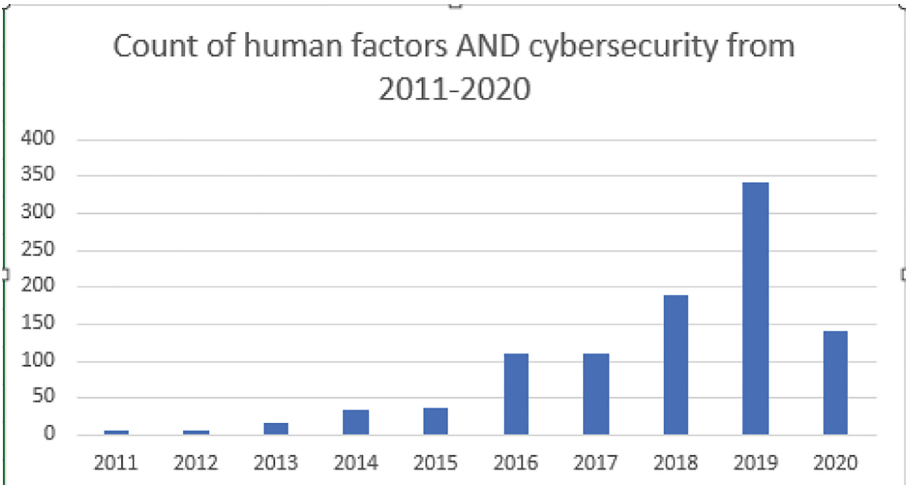

Fig. 2. Trend analysis of articles on cybersecurity and human factors (AuthorMapper, n.d.)

4 Results

4.1 Content Analysis Based on Leading Terms

The bibliometric data from Harzing, which included terms from titles and keywords, was used in VOS Viewer for cluster analysis on the most frequently used and related terms. The process included creating a map based on text data from the Harzing extraction, and setting the parameters for the terms to be selected. The parameter was to have greater than or equal to ten occurrences out of the 2121 terms. From 980 articles in Google Scholar ranging from 2002–2020, 18 terms met the parameters, and all terms were included in the cluster analysis. The clusters are colored based on the average year of occurrence, with the color spectrum key at the bottom right. See Fig. 3.

Table 1 shows the rate of occurrences from the 980 articles from the 19 years. Logically, the top two results are the terms that were searched for initially. Nevertheless, the term that was surprising to find on the list, with the seventh-highest occurrences, was AHFE (Applied Human Factors and Ergonomics). International conference was also on the list, which indicates that the greatest quantity publications on the topic were from the AHFE International Conference. The list of sources from AuthorMapper also supported this assessment with 131 publications in the conference book.

To understand more about AHFE as a leading term in the GoogleScholar search from Harzing, Springer's AuthorMapper is reviewed in more detail under the search term "human factors" AND cybersecurity. AuthorMapper shows that "Advances in Human Factors in Cybersecurity" (from AHFE Conference) is the leading publication in terms of number of articles included as of early 2019. As 'leading publication,' it contains 131 out of 1002 listed articles. It is 1st among 543 publications listed. The following tables show leading authors, years of publication, keywords showing emerging themes emphasized by authors and institutions as well as a count of articles contained in the database on this search topic see Table 2.

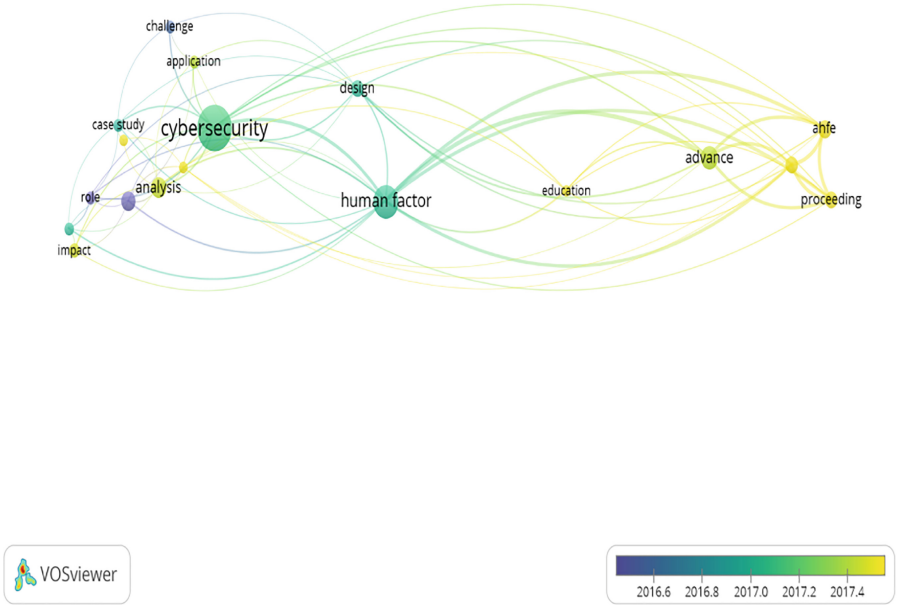

Fig. 3. This figure shows leading terms from cluster analysis in VOS Viewer (Visualization of Similarities). The map is based on metadata captured in Harzing from a search of "human factors" AND "cybersecurity" capturing 980 articles listed within Google Scholar from 2002–2020.

Table 1. Table of leading terms 2002–2020

Term	Occurrences	Relevance
Cybersecurity	236	0.73
Human factor	125	0.09
Advance	56	0.20
Security	44	1.04
Analysis	44	0.94
Design	33	0.05
Ahfe	32	0.33
Proceeding	30	0.33
International conference	29	0.31
Impact	22	1.75
Cyber security	20	1.47
Application	20	1.18
Challenge	19	1.79
Role	19	1.33
Case study	18	1.68
Education	17	0.07
Human	16	1.00
Cybersecurity education	15	3.72

Using the Harzing data from Google Scholar, another cluster map was created for Fig. 4 by the same process as Fig. 3, although with the data from 1000 articles uses the timeframe of 2015–2020. It had several new terms in the map; however the most notable was security. On the other hand, the size and relevance of the AHFE point increased, which indicates that the conference is still emphasizing the importance of research on this emerging area see Table 3.

4.2 Co-citation Analysis

The co-citation analysis is the frequency in which two documents appear together in the reference section of another article (Fahimnia et al. 2015). The articles in the co-citation analysis were taken from two sets of data in the Web of Science database. In the criteria of the search, the parameters were set to only acknowledge articles with three or more co-cited references. The first set of data from 2012–2020 yielded 15 results. The second

Table 2. The table shows leading authors among 2256 listed authors in the AuthorMapper database. Leading keywords show emerging themes emphasized by these leading authors.

Author	Years	Leading keywords	Count
Linkov, Igor	2013–2019	Resilience, Risk, Security, Counterfeiting, Cybernetworks	8
Gonzalez, Cleotilde	2013–2020	Deception, Honeypots, Attack, Behavioral cybersecurity	6
Still, Jeremiah D.	2016–2020	Authentication (graphical and alphanumeric), Cybersecurity, Distorted images	6
Dutt, Varun	2016–2020	Deception, Honeypots, Attack, Behavioral cybersecurity	5
Helkala, Kirsi	2016–2019	Performance, Cognitive ability, Human factors, Socio-technical system	5

Table 3. A table of leading institutions from AuthorMapper shows three countries are represented. Count information is included. Leading keywords show institutional emphasis.

Institution	Country	Leading keywords	Count
University of Oxford	U.K.	Artificial intelligence, Human factors, Privacy, Security, Smart cities	15
Carnegie Mellon University	USA	Deception, Honeypots, Attack, Behavioral cybersecurity, Calibration	14
University of Maryland	USA	Password authentication, Personal data availability, Secondary authentication, User behavior, Anticipatory ethics	11
Old Dominion University	USA	Authentication, Cybersecurity, Graphical authentication, Agent-based modeling and simulation	10
Norwegian University of Science & Technology	Norway	Cyber, Cyber security, Human factors, IoT, Security	8

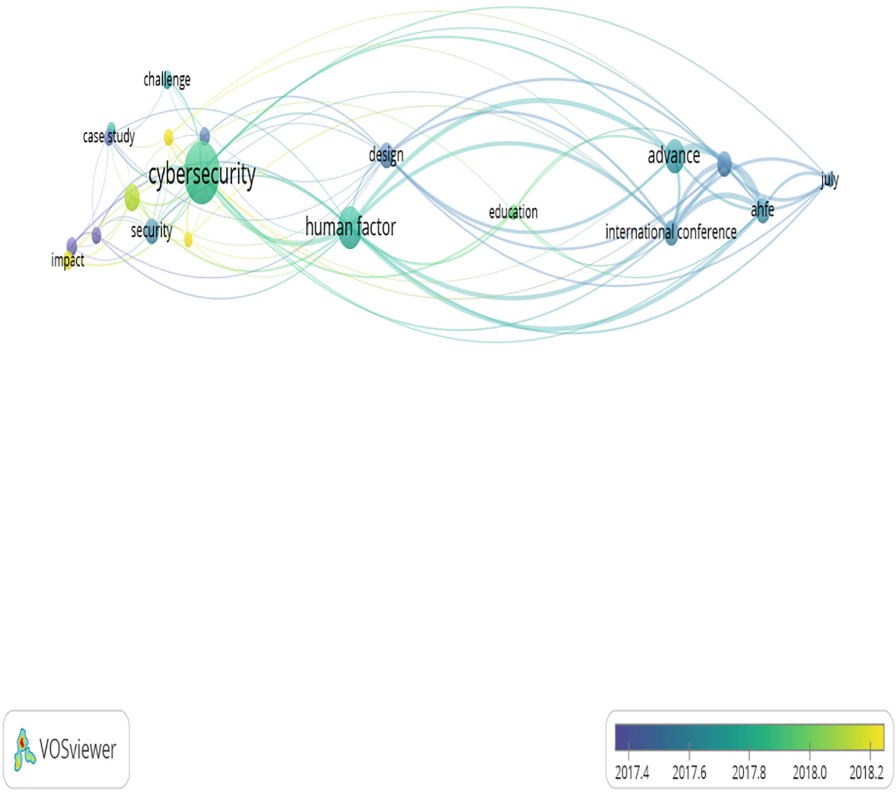

Fig. 4. This figure shows leading terms from cluster analysis in VOS Viewer. The map is based on metadata captured in Harzing from a search of "human factors" AND "cybersecurity" capturing 1000 articles listed within Google Scholar from 2015–2020.

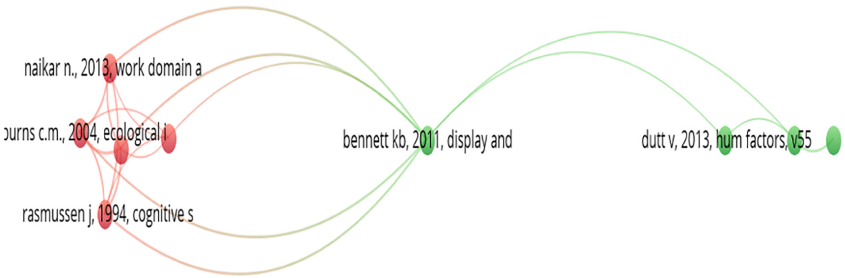

Fig. 5. Co-citation analysis of 48 WoS articles from 2012–2020 and 44 WoS articles from 2015–2020 in VOS Viewer

set of data also yielded 15 results and created an identical cluster map of the co-citation analysis. See Fig. 5 see Table 4.

Dutt is in this co-citation analysis and the table of leading authors in the cybersecurity and human factors field. Rassmussen, Vicente, Flach, and Burns have a foundation in

Table 4. The author, title, publication information, and years from the cluster analysis in Fig. 5.

Authors	Title and Publication Info	Year
Brady, A., N. Naikar, and A. Treadwell	"Organisational storytelling with work domain analysis: Case study of air power doctrine and strategy narrative." In *MODSIM*	2013
Burns, C. M. and Hajdukiewicz JR	Ecological Interface Design, CRC Press	2004, 2017
Vicente, Kim J.	Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press	1999
Rasmussen, Jens, Annelise Mark Pejtersen, and Len P. Goodstein	Cognitive systems engineering. Wiley	1994
Conti, Greg	Security data visualization: graphical techniques for network analysis. No Starch Press	2007
Bennett, Kevin B., and John M. Flach	Display and interface design: Subtle science, exact art. CRC Press	2011
Dutt, Varun, Young-Suk Ahn, and Cleotilde Gonzalez	"Cyber situation awareness: modeling detection of cyber attacks with instance-based learning theory." *Human Factors* 55, no. 3, 605–618	2013
Pattinson, Malcolm, Cate Jerram, Kathryn Parsons, Agata McCormac, and Marcus Butavicius	"Why do some people manage phishing emails better than others?". *Information Management & Computer Security* 20, no. 1: 18–28	2012

Human Factors and Ergonomics. This is further support that authors in the cybersecurity field are impacted by the literature of human factors.

4.3 Content Analysis from MAXQDA

A set of core articles was collected from the extensive collection that was used for bibliometric analysis. They were selected from ResearchGate, Springerlink, IHF Cyber (an Integrated Human Factors Cybersecurity company), and Google Scholar. The co-citation analysis and additional reading lead to the selection of the core articles.

Word Cloud. Articles were imported into MAXQDA to generate a word cloud. The top terms were a combination of human factors terms and cybersecurity terms. The term in the map that stood out in the diagram was "Wickens". Wickens is one of the authors in the *Handbook of Human Factors and Ergonomics.* (Salvendy 1990). His name (word) in the cloud was relatively large, due to the number of times that it appeared within the core articles. See Fig. 6.

Extended Lexical Search. The MAXQDA software is also capable of other literature analyses, such as an extended literature search. In this section of the content analysis, select terms are outlined from among common themes of leading authors or leading terms in the word cloud from maxQDA, the VOS Viewer analysis highlighting leading items within the clusters, and the co-citation analysis within VOS Viewer. Two main categories for the selected terms are "cybersecurity" and "human factors" see Table 5.

Ecological Interface Design: The article (book) in the co-citation analysis by Burns, "Ecological interface design", prompted an extended literature (lexical) search. The

Fig. 6. The Word Cloud for key terms in the core articles (#2–4, 8, 9 13, 15, 22 in the reference section) within the MAXQDA software (project) from the 150 leading terms

Table 5. This table summarizes the terms shown in more detail as part of the content analysis.

Human factors	Cybersecurity
Ecological interface design	Honeypots
Information processing	Risk
Systems design	Privacy

interface design yielded links back to the *Handbook of Human Factors and Ergonomics* (Salvendy, 1990; 2012) and. This connection to the co-citation analysis supports the claim that Human Factors is central to research in human factors with cybersecurity.

Information Processing: Information processing appears in the reference section several times in Moallem's HCI and Cybersecurity Handbook (Moallem 2019). It also appeared in the *Handbook of Human Factors and Ergonomics* (Wickens and Carswell 2012). That section emphasizes capabilities and limitations of people and summarizes how information is selected (attention), processed (perceived) and comprehended (memory aspects).

Systems Design: Systems design is an integral part of the industrial engineering process. When creating a system to protect against a vast array of attacking capabilities, the designer needs to consult knowledgeable sources on systems design. Systems design was

found in the *Handbook of Human Factors and Ergonomics* (Salvendy 1990) in chapter 2 (Czaja and Nair 2012). Emphasis is given to the human factors aspects concerned with interaction of humans with other elements of the system. Beyond the physical aspects, the behavioral aspects are of greater interest in cybersecurity.

Risk: Risk is selected among areas of emphasis of leading author, Igor Linkov, also shown in the table of leading authors from AuthorMapper. Risk relevant to the design process of all cybersecurity technologies. One related article emphasizes comparative risk assessment, recent developments and applications (Linkov et el. 2006). A human factors model that directly relates to the risk assessment is emphasized in Reason's research. James Reason's Swiss Cheese model models a defense against failure with a statistical calculation for the probability of failure, based on the layers of defense with assumptions that one layer may be breached while the next may not (Reason 2006). Redundancy, resilience and human reliability arise as terminologies for further consideration of risk and risk analysis.

Honeypots: This term honeypot is useful as an example that shows how human factors theory can be used in support of cybersecurity. A honeypot is a cybersecurity item that deceives a cyber-attacker into targeting it. When the honeypot is targeted for the attack, it is assumed to have highly sought data and whatnot. Then the attacker is quarantined and loses access to the system. It uses human limitations and desires to tempt the attacker to fall into the trap. The term honeypot is selected among areas of emphasis of leading author Varun Dutt, shown in the table of leading authors from AuthorMapper. The work of Dutt is also referred to among leading articles in the co-citation analysis. That article from the co-citation analysis was published in the Human factors journal (Dutt et al. 2013).

Privacy: Privacy is a leading term in the word cloud and is sometimes considered together with security and trust. Some researchers have emphasized privacy compliance as well as privacy-preserving and privacy-increasing technologies (Moallem 2019). Cybersecurity and protection of privacy, many times, are considered together. The idea of proactive security measures for prevention is preferred to the consequences of loss of reputation and the administrative requirements to notify after a breach of security or privacy.

5 Conclusions and Future Work

Cybersecurity is a rapidly changing field, which involves solving high tech problems with logical defenses. Human factors is steadily gaining recognition in research for cybersecurity systems. Due to the results of the co-citation analysis, it is clear that many of the leading authors recognize the importance of human factors in systems design. The trend analysis shows the increasing awareness and number of publications on the topic, which will help cybersecurity continue to grow and flourish.

Examples of recent funded work from the National Science Foundation in the USA highlight aspects of human-automation interaction and consider practical applications

of privacy and security. The proposal awarded to Patricia Delucia and James Yang of Texas Tech in 2016 is titled Translational Research in Psychological Sciences. The work emphasizes research experience for undergraduates and expects both scientific and societal benefits including applications in cyber-security. Their proposal is intended to support training for a growing demand for human factors professionals.

The research award of DeLucia and Yang is intended to advance research with implications for behavior intended to reduce traffic crashes, improve patient safety and inform human-robot interaction in the context of social robots. One article that was produced as a result of the research so far addresses robots that take on human characteristics. Research related to anthropomorphism may be of interest to the reader as part of future work related to cybersecurity and human factors. Anthropomorphism is the term for computing and/or automation that takes on human characteristics. The publication related to the theoretical and practical implications for anthropomorphism research was published in an ACM/IEEE conference on human-robot interaction in 2018 (Jones 2018). Additional information about related projects can be found at NSF.gov using the keywords "cyber security" and "human factors" in search.

References

AuthorMapper. https://www.authormapper.com/. Accessed 03 Jan 2020

Czaja, S.J., Sankaran, N.N.: Human factors engineering and systems design. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 4th edn, Chap. 2, pp. 38–54. Wiley, New Jersey (2012)

Dutt, V., Ahn, Y.-S., Gonzalez, C.: Cyber situation awareness: modeling detection of cyber attacks with instance-based learning theory. Hum. Factors **55**(3), 605–618 (2013)

Fahimnia, B., Sarkis, J., Davarzani, H.: Green supply chain management: a review and bibliometric analysis, 23 January 2015. https://www.sciencedirect.com/science/article/abs/pii/S0925527315000067

Google Scholar. https://scholar.google.com/. Accessed 03 Jan 2020

Jones, K.S., Niichel, M.K., Armstrong, M.E.: Robots exhibit human characteristics: theoretical and practical implications for anthropomorphism research. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, p. 137 (2018)

Harzing's Publish or Perish. https://harzing.com/resources/publish-or-perish. Accessed 03 Jan 2020

IHF Cyber: Cyber Security : A Human Factors Dichotomy. N.p. Print

Linkov, I., Satterstrom, F.K., Kiker, G., Batchelor, C., Bridges, T., Ferguson, E.: From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. Environ. Int. **32**(8), 1072–1093 (2006)

MAXQDA. https://www.maxqda.com/. Accessed 03 Jan 2020

Mendeley. https://www.mendeley.com/?interaction_required=true. Accessed 03 Jan 2020

MIT: 23 faculty members awarded tenure, 25 October 2006. http://news.mit.edu/2006/tenure-1025

Moallem, A. (ed.): Human-Computer Interaction and Cybersecurity Handbook. CRC Press, Boca Raton (2019)

NAE: Professor Robert Morris (2019). https://www.nae.edu/204131/Professor-Robert-Morris

Reason, J., Hollnagel, E., Paries, J.: Revisiting the Swiss cheese model of accidents. J. Clin. Eng. **27**(4), 110–115 (2006)

ResearchGate. https://www.researchgate.net/. Accessed 03 Jan 2020

SpringerLink. https://link.springer.com. Accessed 03 Jan 2020
United States v. Robert Tappan Morris, 928 F.2d 504. https://www.courtlistener.com/opinion/557785/united-states-v-robert-tappan-morris/
VOSviewer. https://www.vosviewer.com/. Accessed 03 Jan 2020
Web of Science. https://apps.webofknowledge.com. Accessed 03 Jan 2020
Wickens, C.D., Carswell, C.M.: Chapter 5 information processing 2 three approaches to information. In: Salvendy, G. (Ed.) Handbook of Human Factors and Ergonomics, 4th Edn, pp. 117–151 Wiley: New Jersey (2012)