Technological Forecasting & Social Change 188 (2023) 122316

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Measuring security development in information technologies: A scientometric framework using arXiv e-prints

Dimitri Percia David [a,b,e,*], Loïc Maréchal [b,d], William Lacube [b], Sébastien Gillard [a,c], Michael Tsesmelis [b], Thomas Maillart [a], Alain Mermoud [b]

[a] Information Science Institute, Geneva School of Economics & Management, University of Geneva, 40 Boulevard du Pont-d'Arve, Geneva 1211, Switzerland
[b] Cyber-Defence Campus, armasuisse Science and Technology, Feuerwerkstrasse 39, Thun 3602, Switzerland
[c] Chair of Defense Economics, Military Academy at ETH Zurich, Kaserne Reppischtal, Birmensdorf 8903, Switzerland
[d] Department of Information Systems, HEC Lausanne, University of Lausanne, Internef Building, UNIL-Chamberonne, Lausanne 1015, Switzerland
[e] Institute of Entrepreneurship & Management, University of Applied Sciences of Western Switzerland (HES-SO Valais-Wallis), Techno-Pôle 1, Le Foyer, Sierre 3960, Switzerland

ARTICLE INFO

Dataset link: arXiv dataset (Original data)

Keywords:
Technology mining
Technology management
Security economics
Time-series econometrics
Natural language processing
Opinion mining

ABSTRACT

We study security-development patterns in computer-science technologies through (i) the security attention among technologies, (ii) the relation between technological change and security development, and (iii) the effect of opinion on security development. We perform a scientometric analysis on `arXiv` e-prints ($n = 340,569$) related to 20 computer-science technology categories. Our contribution is threefold. First, we characterize both processes of technological change and security development: while most technologies follow a logistic-growth process, the security development follows an AR(1) process or a random walk with positive drift. Moreover, over the lifetime of computer-science technologies, the security development surges at a late stage. Second, we document no relation between the technological change and the security development. Third, we identify an inverse relation between security attention and experts' opinion. Along with these results, we introduce new methods for modeling security-development patterns for broader sets of technologies.

1. Introduction

In this paper, we investigate the security-development process of computer-science technologies. We provide the first systematic and quantitative investigation of security development and its ramifications for technological change and opinion formation. We aim to respond to both academic and practical needs for understanding the security development of information systems and its relation to social change.

Information technologies increasingly impact societies and political environments by allowing for real-time communication, social media interactions, or management of critical infrastructures. For instance, the democratic-debate shifts towards social networks and law enforcement rely on computer vision and machine learning to identify and even predict crime. Furthermore, information technology is now an essential determinant in warfare and the sovereignty of nations. Information technologies are also prone to security failures affecting the integrity, availability, and confidentiality of data. These failures impact individuals and organizations alike. The recent ransomware attack on the Colonial pipeline in the United States is a direct reminder of the fragility of our digital societies and explains the increasing popularity of the digital trust concept (Tsvetanov and Slaria, 2021).

Technological change redefines information and communication technologies (ICTs) (Shalf, 2020). For organizations, emerging technologies carry both opportunities to enhance operations efficiency for organizations (Brock, 2021) and security threats (Laube and Böhme, 2017; Anderson and Moore, 2006; Jang-Jaccard and Nepal, 2014). Overcoming these threats requires the implementation of methods such as secure-by-design (SBD) engineering, which considers security from the first stage of the technology development (Anderson, 2020) and is similar to probabilistic safety assessments (PSA) common in the aerospace and nuclear industries. Yet, the literature in security economics affirms that information security is often absent in early

* Corresponding author at: Information Science Institute, Geneva School of Economics & Management, University of Geneva, 40 Boulevard du Pont-d'Arve, Geneva 1211, Switzerland.
E-mail address: dimitri.perciadavid@hevs.ch (D. Percia David).

[1] We use the term "security" as a replacement for "information security" throughout the text. We define the term as the practice of protecting the privacy, integrity, availability, and non-repudiation of data. Thus, information security is the probabilistic reduction of unauthorized/inappropriate data access and unlawful use, disclosure, disruption, deletion, corruption, modification, inspection, recording, or devaluation of information (Anderson, 2020).

https://doi.org/10.1016/j.techfore.2023.122316
Received 5 March 2022; Received in revised form 1 November 2022; Accepted 2 January 2023
Available online 11 January 2023

stages of technological change (Böhme, 2013) given its high cost, engineering complexity, and the misalignment of priorities between end-users and security providers (Anderson and Moore, 2006; Böhme, 2013; Anderson, 2020).[1]

While there exists a rich and growing literature on the limited attention paid to security in information technologies, there is a dearth of research on the actual evolution of security along the development of technologies, a concept that we coin *security development*. In this paper, we investigate three aspects of security development: (i) the evolution of the security attention embedded in technologies, (ii) the relationship between technologies and security developments, and (iii) the effect of the opinion towards security development. We adopt a technology-mining (scientometric) approach on 1 854 076 e-prints of the arXiv open-data repository (from August 14, 1991, until December 31, 2020). Out of this sample, we identify 340 569 e-prints related to 20 *Computer-Science Technology Categories related to Cybersecurity* (hereafter, CSTCCs), on which we examine the three aspects of security dynamics mentioned above. We aim to build the first indicator of security development in computer-science technologies.

Our contribution is threefold. First, we find strong support for the view that technological change follows a common logistic growth process when we allow the parameters to be idiosyncratic. We also characterize the security development of technologies and find that its proxy variable, the security attention, follows either a simple first order autoregressive process or a random walk with a positive drift. For both processes, the positive trend in security attention and additional results support the view that security is taken into account only at later stages of technological change. Second, we are not able to identify a significant relation between the technological change and the security development in both absolute levels or growth rates. Third, we identify determinants of the security development approximated by a measure of security attention. As determinants, we use both the opinion extracted from scientific articles as well as the dispersion of this opinion across the lifetime of a CSTCC. We identify a significantly negative (positive) relation when the opinion (dispersion of opinion) is used as explanatory variables.

These results shed light on the dynamics of security development, which is essential to understand (i) how security evolves, (ii) the relationship between technological change and security development, and (iii) the determinants of security development. In a further step, we discuss how our results may be used to shape guidelines and principles. More specifically, we suggest using the quantitative evaluation performed in this work to create benchmarks and avoid weak links (*i.e.*, technologies which perform poorly). Additionally, such quantitative evaluations help estimate the level and dynamics of open security among technologies, an essential security factor promoted by the NIST cybersecurity framework.

The remainder of this article proceeds as follows. Section 2 presents the literature review and hypotheses. Section 3 details the data and the methodology. Section 4 presents the results. Section 5 discusses the implications and limitations. Section 6 concludes.

2. Literature review, theory and hypotheses development

In this section, we review the methods developed for measuring the theoretical variables related to (i) technological change, (ii) security development, and (iii) the relation between *opinion* and security. We emphasize the research gaps that we exploit in this work to offer pertinent findings, hence giving relevancy to our approach. Such gaps are the scarcity of (i) benchmarking indicators related to technological change within an inter-technologies context, (ii) holistic indicators of security development, and (iii) research on how security attention is associated with *opinion*. Given these gaps, our hypotheses question the existence of three potential patterns: (i) the dynamics of *security attention*, (ii) the relationship dynamics between technological change and security development, and (iii) the dynamics between *opinion* and security attention. We investigate these gaps within *Computer-Science Technology Categories related to Cybersecurity* (CSTCCs).

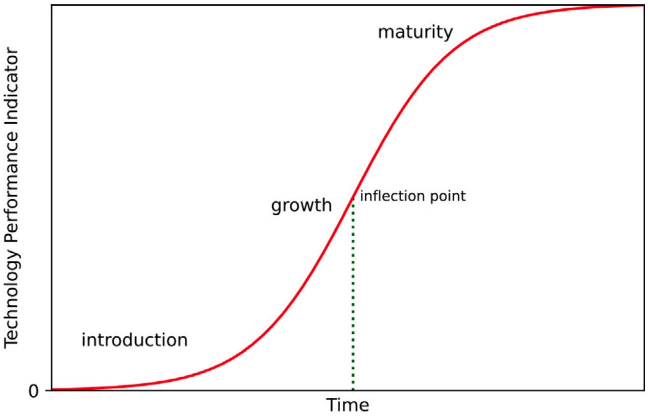

Fig. 1. This figure shows the logistic growth process of technological change. The y-axis is a performance indicator. Within this process, an inflection point (projected on the x-axis, in green) is reached, and corresponds to the moment from which the marginal performance decreases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. Technological change and security development processes

Both qualitative and quantitative approaches have been used to model technological change and characterize its process, common to all technologies. Qualitative assessments include, in particular, Schumpeter (1942) and Jaffe, Newel, and Stavins (2002) who consider three consecutive discrete stages: invention, innovation, and diffusion (Jaffe et al., 2002). Rogers (2010) specifies the invention stage as the engineering inception of a technology, the innovation stage as its practical implementation, and the diffusion stage as its commercialization (Rogers, 2010).

Quantitative methods on the other hand assume a functional form for the technological change process, which allows for studies in continuous time. The sigmoid function, a special case of the logistic function, stands out in the literature (for an extensive literature review, see Coccia (2005), Lee (2021), Haleem et al. (2019), Calleja-Sanz et al. (2020), Daim et al. (2016), Porter et al. (2011)). Interestingly, this continuous-time growth process can be still broken down into three stages: an explosive stage in the middle of the life cycle surrounded by two soft ones at the beginning and at the end. These findings reflect those uncovered qualitatively by *e.g.* Jaffe et al. (2002). These quantitative studies however use a different nomenclature and define these stages as introduction, growth, and maturity (Li et al., 2019b; Rogers, 1995, 2010; Lotfi et al., 2014; Chen et al., 2011; Adamuthe et al., 2014; Andersen, 1999; Priestley et al., 2020). The introduction stage typically consists of a slow positive change rate, though this rate is highly specific to the technology category and its environment. Years of gestation may be needed before an emerging technology achieves widespread market acceptance and commercial success. By satisfying the users' needs of niche markets, a novel technology improves before reaching a broader population of mainstream users (Adner and Levinthal, 2002). During the subsequent growth phase, the mass market adopts the technology, which becomes compelling, attracts more investments, and sees overall improvements thanks to new features (Klepper, 1997; Perez, 2010). Last, the positive change rate decreases as the technology matures and approaches its diffusion limits due to market saturation (Perez et al., 2010; Abernathy et al., 1978).

The empirical literature of these quantitative approaches uses different private- and open-source datasets to measure technological change. Some studies rely on indicators such as bibliometric analyses related to scientific publications (*i.e.*, scientometrics) (Zhang et al., 2020b,a; Mayr et al., 2014; Dotsika and Watkins, 2017; Jaewoo and Woonsun, 2014; Rezaeian et al., 2017), patents (Hajikhani and Suominen, 2022; Choi et al., 2022; Chen et al., 2017; Golembiewski et al., 2015; Noh et al., 2016; An et al., 2018; Lee and Lee, 2019; Song et al., 2017), industry-market indicators (job openings, trade registers, networks of authors and citations), or a combination of these (An et al., 2022; Xi et al., 2022; Ali et al., 2022; Daim et al., 2012; Lee et al., 2010). To analyze this data, text-mining methods (Chen et al., 2017; Choi and Hwang, 2014; Hao et al., 2014; Antons et al., 2020) and network analysis (Chen et al., 2022; Zhang et al., 2022; Li et al., 2019b; Mikheev, 2020; Feng et al., 2022; Hong et al., 2021) are particularly prolific. For instance, Guo, Wang, Tian, and Xian (2019) analyze 1666 scientific publications to detect trends in network technologies and information systems (Guo et al., 2019). With a literature-growth approach and a co-citation analysis, they detect exponential patterns (capturing the introduction and the initial stages of the sigmoid process) in technologies. Similarly, Priestley, Sluckin, and Tiropanis (2020) investigate a longitudinal dataset of 20493 Internet-related US patents between 1990 and 2013. They find that the accumulation of corporate Internet inventions follows a sigmoid growth process (Priestley et al., 2020). Son, Kim, Kim, Han, and Kim (2010) extract topics and primary contributors of peer-reviewed publications. Using network methods, they extract global trends in automation and robotics technologies (Son et al., 2010). Similarly, technological change has been investigated in fields other than information security (*e.g.*, Chen et al. (2011), Bengisu and Nekhili (2006), Lotfi et al. (2014), Erzurumlu and Pachamanova (2020)). The vast majority of these approaches have been employed for individual technologies (Daim et al., 2016). These analyses uncover a common change process to all technologies and estimate the technology-specific parameters of this process.

Rogers (2010) states that technological diffusion occurs within a social system (Rogers, 2010). In this respect, the `arXiv` community is a typical social system in which e-prints related to technologies are communicated through uploads on a shared repository. Furthermore, in many technical fields such as mathematics, physics, and computer science, an important share of research papers are self-archived on the `arXiv` repository before being submitted and subsequently presented or published either at a conference or in a peer-reviewed journal (Sutton and Gong, 2017). Therefore, we assume `arXiv` to be a representative source of observation for technological changes. The monthly number of uploads of e-prints related to CSTCCs can be used to assert whether technological change follows a common (logistic growth) process. We state our first null hypothesis:

H1a: *There is no common process that characterizes the technological change of CSTCCs.*

While quantitative methods are widely used to analyze technological change, they have seldom been used in the analysis of security development. One evident exception is the information security field itself, which uses indicators built upon skill indices (Carlton, 2016), organizational development (Goode et al., 2018), job openings (Assante and Tobey, 2011), risks (Hubbard and Seiersen, 2016), dynamics of incidents (Liu et al., 2015) and evolution of behaviors (Li et al., 2019a). For a systematic literature review, see Meland et al. (2021). This research has led to the development of good practices and regulation in the information security industry. A notable example in the field is the security by design (SBD) principle, which consists in embedding security attention within the technological development process (Santos et al., 2017; Casola et al., 2020; Kreitz, 2019). Under SBD, the security attention of information systems is designed in parallel from the first stage (Casola et al., 2020). However, as stated above, there is a lot of pressure to commercialize a product quickly for it to become profitable as fast as possible, which in turn postpones the security attention to a later stage of development (Böhme, 2013; Anderson and Moore, 2006; Anderson, 2020). Therefore, our null hypothesis is:

H1b: *The security development of CSTCCs is time-invariant.*

2.2. Misalignment between technological change and security development processes

Among all technologies, those related to computer-science are subject to a particularly high number of stages, many of which are related to engineering and project management work. These stages start from the analysis of usability needs and end at the deployment of the technology (Zharov and Kozlov, 2018). In this life-cycle, the security is evaluated, designed, and implemented (Howard and Lipner, 2006) at various stages, again with large variations from one technology to the other (Anderson, 2020). Previous research points out discrepancies between drivers of technological change and those of security development. For instance, research in information-security economics highlights the fact that misaligned incentives between developers and end-users constitute a significant barrier in the security development of a technology (*e.g.*, Anderson and Moore (2006), Anderson (2020, 2001), Anderson and Moore (2007), Böhme (2013)). Notably, Anderson and Moore (2006) state that, for the security development of large-scale systems, incentives matter at least as much as technical aspects. Security failures arise when individuals in charge of a technology's security are not the ones who suffer the costs of failures (Anderson and Moore, 2006). Thus, this misalignment of incentives detaches the security development of CSTCCs from their technological change. Similarly, Anderson (2007) points out that software markets behave in ways akin to a *market for lemons* (Anderson and Moore, 2007). Following the concept of Akerlof (1970) (Akerlof, 1978), he examines how the quality of security in the software market degrades in the presence of information asymmetry between buyers and sellers. In both security and general software markets, most users cannot assess the vulnerability status of the products they purchase. Thus, the buyers' incentives to pay for security shrink, thereby reducing the sellers' willingness to improve security. This mechanism leads to a detachment of security development from other technological change aspects. Finally, the urge to improve and uphold business revenue streams by launching new products before competitors tends to undermine the security development of a technology. Sellers often commercialize technologies with underdeveloped security and use customers as implicit testers to identify and patch vulnerabilities (Anderson, 2020). Moreover, the security attention is often skipped in the early stages of technological change given the substantial investments security requires Anderson and Moore (2006), Böhme (2013). We state our null hypothesis:

H2: *The security development of a CSTCC is independent from the technological change.*

2.3. Factors of security development

Trend analysis typically captures the attention of a community towards a technology (Daim and Yalçin, 2022). However, it does not capture the opinion (Jun et al., 2012). Yet, such opinion is an essential indicator of technological efficiency (Liu, 2012). The academic literature on a topic is written by experts (engineers and scholars), who use a lexicon that carries information about their opinion towards a technology. Capturing opinion in unstructured data is possible thanks to sentiment analysis, which is defined as the analysis of a lexicon of texts transformed into structured data using natural language processing (NLP) (Liu, 2012). The method is widely used in management, finance (Liu, 2012; Fang and Zhan, 2015; Chang and Wang, 2018; Maks and Vossen, 2013) and marketing to predict consumer trends (Bai, 2011). Even though researchers are supposed to be opinion-neutral, it is likely unavoidable for their texts to carry a latent opinion. Furthermore, research has investigated the interaction between opinion and security attention (Pletea et al., 2014). Gurung and Raja (2016) show that the prevalence of privacy and security aspects affects individuals' risk perceptions (Gurung and Raja, 2016). The connection between risk and opinion is additionally widely studied in several fields (Chang and Wang, 2018). Yang et al. (2015, 2016) use sentiment analysis to show the relation between consensus and technological uncertainty affecting perceived risk (Yang et al., 2016, 2015). Therefore, the prevalence of security attention of experts on a CSTCC should be positively related to the opinion. Our null hypothesis is:

H3a: *For each CSTCC, the experts' opinion does not explain the security development.*

Both product engineering and scientific works are iterative processes requiring design thinking and peer-reviewing, respectively (Steinmetz, 2011). This process ends once a consensus is reached. For example, Dou, Zhang, and Nan, (2017) and Lehrer and Wagner (2012) show that the opinion related to a product converges to a consensus along with the product's improvement (Dou et al., 2017; Lehrer and Wagner, 2012). Similarly, Yüzügüllü and Deason (2007) show that the maturity and market-readiness of a technology are factors of consensus in the community surrounding it (Yüzügüllü and Deason, 2007). In addition, the consensus is easy to estimate from measures of opinion, for instance using the cross-sectional standard deviation of sentiments at one point in time (Huang et al., 2019). We expect to observe similar evidence in any technology field in general and in scientific works related to technology development in particular. Consequently our related null hypothesis is:

H3b: *For each CSTCC, the experts' consensus does not explain the security development.*

3. Data and methods

In this section, we first present the empirical variables and how we measure them. These empirical variables (the e-prints, the security attention, and the opinion) are the ones we use as proxies for the latent variables (the technological change, the security development, and the opinion, respectively). We then present the methodologies we use to test our hypotheses.

3.1. Data

We extract data from open scientific works (*i.e.*, scholar articles consisting of working papers, preprints, technical reports, post-proceedings, and publications) labeled e-prints and uploaded on the `arXiv` repository. The latter is a free distribution service and open-access archive for academic articles related to various technical fields, including computer science (uploaded e-prints are not peer-reviewed). First, we download the entire `arXiv` repository (1 858 293 files, corresponding to 3 TB of text in pdf) through a mirror of the database found on `kaggle`.[2] The data encompasses all e-prints uploaded since the inception of the `arXiv` repository (August 14, 1991) until December 31, 2020. Next, for each CSTCC, we (i) count the number of e-prints through time, (ii) extract the share of e-prints that include security attention, and (iii) extract the opinion expressed by authors.

3.1.1. e-prints

We construct our main variables from the set of e-prints. We use the number of articles in each CSTCC as a proxy for the technological change and the text itself to capture the security attention and opinion.

To consistently classify and archive all e-prints, `arXiv` representatives (composed of a scientific advisory board) have a systematic category taxonomy.[3] They determine this taxonomy with a Delphi-like method involving expert members for each `arXiv` scientific field.[4] This implies that authors willing to upload their e-prints on `arXiv` must select the corresponding category. Then, `arXiv` moderators check the authors' classification to ensure consistency. We consider this three-step classification to be robust because (i) the taxonomy is created through a consensus reached by a panel of experts, (ii) authors have no apparent incentive to misclassify their work, and (iii) moderators check the classification consistency. As e-prints are attached to various predetermined `arXiv` fields unrelated to computer science (such as physics, mathematics, quantitative biology, quantitative finance, and economics), we filter the `arXiv` predetermined fields to extract computer-science technologies (denoted `cs.`) repository. We apply a second filter, considering `arXiv` subcategories in the `cs.` fields that are directly associated with information-security technologies. To determine which `arXiv` subcategories of the `cs.` repository are effectively related to information-security technologies, we use the *Defenses* sections listed in the Information Security portal of Wikipedia as a reference.[5] Thus, we select the `arXiv` subcategories of the `cs.` repository whenever this subcategory is also mentioned within the Wikipedia Defenses section.

From this two-step selection procedure, we retain 20 subcategories of CSTCCs. In the case of the `cs.` repository, the category taxonomy substantially relies on the list of methodology and technology categories provided by the *2012 ACM Computing Classification System*.[6] Therefore, we consider the 20 categories mentioned above as distinct CSTCCs. We depict the list of these CSTCCs and their respective number of e-prints in Table 1.

If the `arXiv` repository is nowadays regarded as an established platform amongst various scientific communities for uploading their e-prints, it enjoyed no such popularity at its inception. Therefore, we cannot assume that the `arXiv` platform depicts a constant attention rate related to each CSTCC. To circumvent this bias, we normalize the number of e-prints related to each CSTCC by dividing the total number of e-prints per CSTCC per period (month) by the corresponding amount of total e-prints (*i.e.*, including all categories) of the `arXiv` repository per period (month). Such a measure is depicted in Figs. 2 and 3. A preliminary analysis shows that, for the great majority of categories, we either witness an exponential trend, depicted in Fig. 2 (*i.e.*, corresponding to the introduction and growth stages of the logistic growth, or an actual logistic growth process, depicted in Fig. 3 (*i.e.*, corresponding to the three stages of the process).

3.1.2. Security attention

The security attention measure relates to (i) the technology's dependability in terms of privacy-preserving and confidentiality aspects, and (ii) the technology's ability to ensure the integrity, availability, and non-repudiation of data. To capture the security attention expressed in e-prints, we thus select a set of keywords related to the two concepts mentioned above. These relate to the well-known CIA triad (*i.e.*, *confidentiality*, *integrity*, and *availability*) and the *non-repudiation* principle (Cherdantseva and Hilton, 2013; Ritzdorf et al., 2017), and these are further explained in the *Information Security* portal of *Wikipedia*.[7] The list of keywords is: *secure*, *security*, *safe*, *reliability*, *dependability*, *confidential*, *confidentiality*, *integrity*, *availability*, *defense*, *defence*, *defensive*, and *privacy*. We then query the `arXiv API` to select e-prints that contain these keywords in either their title or abstract. Subsequently, to extract the share of security attention among each CSTCC, we divide the number of e-prints per CSTCC including these keywords by the total number of e-prints per CSTCC. Fig. 4 depicts how the share of e-prints alluding to security has changed, illustrated by the CSTCC *computer vision and pattern recognition*. A preliminary analysis shows that, for the great majority of categories, we witness (i) a diminishing dispersion and (ii) an upward trend of the measure.

[2] https://www.kaggle.com/Cornell-University/arxiv

[3] https://arxiv.org/about/people/scientific_ad_board#advisory_committees

[4] https://arxiv.org/category_taxonomy

[5] https://en.wikipedia.org/wiki/Information_security

[6] https://arxiv.org/corr/subjectclasses

[7] https://en.wikipedia.org/wiki/Information_security#Key_concepts

Table 1
arXiv categories (corresponding CSTCCs) and their respective count of e-prints, with and without security attention. The category cs.CR (*Cryptography and Security*) has a share of security attention greater than 75%.

arXiv categories	Cluster name (CSTCC)	Total count of e-prints	With security attention	% of security attention
cs.AI	Artificial Intelligence	38 620	11 447	29.640
cs.AR	Hardware Architecture	2 573	971	37.738
cs.CC	Computational Complexity	8 492	1 216	14.319
cs.CL	Computation and Language	29 528	8 536	28.908
cs.CR	Cryptography and Security	19 784	14 952	75.576
cs.CV	Computer Vision and Pattern Recognition	64 696	21 852	33.776
cs.DB	Databases	6 269	2 341	37.342
cs.DC	Distributed, Parallel, and Cluster Computing	14 955	5 686	38.021
cs.DS	Data Structures and Algorithms	18 269	3 458	18.928
cs.GT	Computer Science and Game Theory	7 992	2 279	28.516
cs.HC	Human–Computer Interaction	8 774	2 753	31.377
cs.IR	Information Retrieval	10 407	3 216	30.902
cs.LG	Machine Learning	94 024	30 142	32.058
cs.NE	Neural and Evolutionary Computing	10 155	2 649	26.086
cs.NI	Networking and Internet Architecture	16 606	6 826	41.106
cs.OS	Operating Systems	652	303	46.472
cs.PL	Programming Languages	5 731	1 937	33.799
cs.RO	Robotics	16 187	6 055	37.407
cs.SE	Software Engineering	10 032	4 109	40.959
cs.SY	Systems and Control	18 347	6 845	37.309

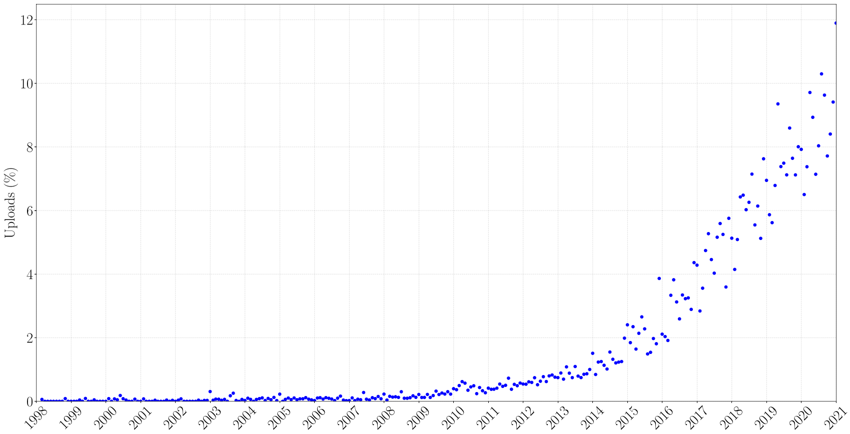

Fig. 2. Normalized count of e-prints: computer vision and pattern recognition. This figure depicts the number of e-prints that belong to the field of "computer vision and pattern recognition" over each month. The data is normalized and spans the period 1998–2021. The figure pictures the "introduction" and "growth" stages of the logistic growth function (see Fig. 1).

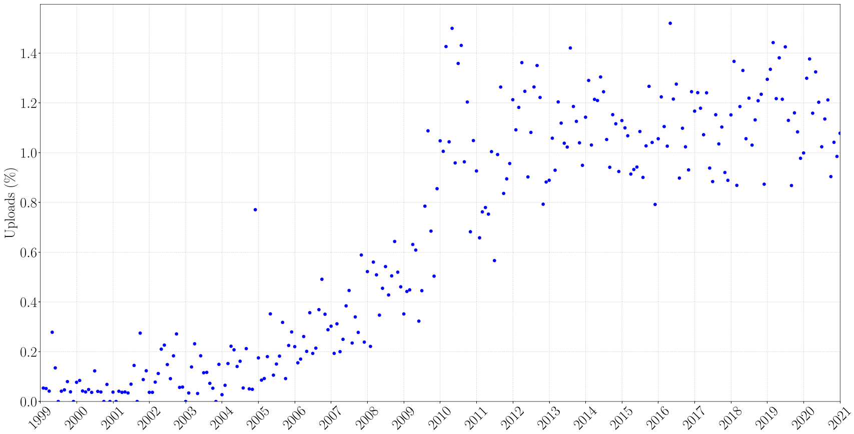

Fig. 3. Normalized count of e-prints: networking and internet architecture. This figure depicts the number of e-prints that belong to the field of "networking and internet architecture" over each month. The data is normalized and spans the period 1999–2021. The figure pictures the three phases of the logistic growth: "introduction", "growth", and "maturity" stages (see Fig. 1).

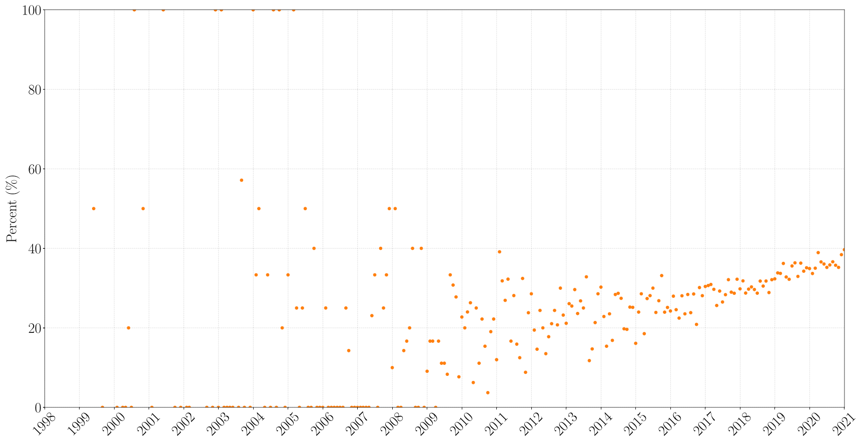

Fig. 4. Security attention: computer vision and pattern recognition. This figure depicts the evolution of the prevalence of security attention, *i.e.*, the e-prints containing security attention keywords divided by the total number of e-prints. The frequency is monthly and the period is 1999–2021. The right-hand side depicts a diminishing dispersion and an upward trend. This pattern is present in the majority of categories, as depicted in the multi-plot of all security attention measures (for each CSTCC) available in Fig. 11 (Appendix). Descriptive statistics of the security attention measures are available upon request.

3.1.3. *Opinion*

To capture the *opinion* of authors related to each e-print attached to a given CSTCC, we employ opinion mining by implementing a classic lexicon-based approach based on a labeled thesaurus (the NLTK opinion lexicon of `Python`, in English) to classify the lexicon of e-print authors as either positive or negative (Serrano-Guerrero et al., 2015).

We first clean and normalize every word in e-prints before transforming them into tokens (machine-readable inputs). Cleaned tokens are obtained through standard NLP procedures such as (i) transforming all text in British English, (ii) removing special characters, stop words, punctuation, and lowering upper-cases. Then, we normalize cleaned tokens through lemmatization (morphological analysis to transform tokens into their canonical form). We also consider the impact of direct quotations referencing previous literature in our sample articles.[8] We consider this problem to be marginal in this analysis. First, because the amount of direct quotations in the scientific literature is very limited and we check this assumption manually by checking a sub-sample of articles of each category. Second, even in the case where these quotations would be significant, they would only generate an error, which would disappear in the aggregation process and not a systematic bias.

Subsequently, we apply a standard cumulative-sentiment function that classifies each token into either a positive or negative opinion before summing the result for each e-print. The final opinion score is normalized for each e-print and ranges from −1 (worst) to 1 (best). For each month, we sum the scores of e-prints related to the same CSTCC.[9] We obtain the opinion distribution for each CSTCC and each month. In Fig. 5 we display an example of the evolution of the opinion for the CSTCC *computer vision and pattern recognition*. We present the corresponding descriptive statistics in Table 2.

Table 2
Descriptive statistics: monthly opinion.

`arXiv` categories	Mean	Median	Std Dev	Skewness	Kurtosis
`cs.AI`	−0.002	−0.001	0.006	−1.341	2.809
`cs.AR`	−0.001	0.000	0.007	−1.592	7.664
`cs.CC`	−0.009	−0.009	0.005	−0.967	3.658
`cs.CL`	0.004	0.005	0.003	−1.038	3.462
`cs.CR`	−0.006	−0.005	0.014	−10.379	140.804
`cs.CV`	−0.003	−0.002	0.007	−2.053	7.548
`cs.DB`	0.001	0.001	0.006	−0.077	9.936
`cs.DC`	−0.001	0.000	0.005	−1.570	11.395
`cs.DS`	−0.004	−0.003	0.005	−0.008	9.114
`cs.GT`	0.002	0.003	0.008	−1.379	14.532
`cs.HC`	0.004	0.005	0.007	−1.261	5.772
`cs.IR`	0.006	0.007	0.007	−3.640	22.239
`cs.LG`	−0.002	−0.001	0.006	−1.737	10.354
`cs.NE`	−0.003	−0.001	0.007	−2.538	13.560
`cs.NI`	−0.002	−0.001	0.005	−1.774	13.112
`cs.OS`	−0.003	−0.001	0.010	−1.441	4.284
`cs.PL`	0.002	0.002	0.006	−3.476	40.685
`cs.RO`	−0.003	−0.002	0.007	−3.337	18.244
`cs.SE`	−0.001	0.000	0.006	−1.022	5.650
`cs.SY`	−0.005	−0.005	0.004	−0.529	7.064

This table displays summary statistics of the monthly opinion for each CSTCC.

3.2. *Methods*

The following subsection presents the methodologies employed to test our hypotheses. For all methods, we define a set Ω_x for all CSTCC, x:

$$\Omega_x = \left\{ t \mid t \leq N_x,\, t \in \mathbb{N}^* \right\} \tag{1}$$

where N_x is the number of months comprised between the first and the last e-print for x.

3.2.1. *Technological change process*

To model the technological change, we use a non-linear optimizer, the *Levenberg–Marquardt* algorithm, to fit a logistic function on the historical observations of e-prints for each CSTCC) (Moré, 1978). We use the `Python scipy` package and its `.optimize.curve_fit` method.[10] We make this choice because the logistic function generalizes the sigmoid and therefore offers more flexibility to capture the common process underlying each CSTCC. It is also a bounded and differentiable function with a single inflection point (Han and Moraga, 1995). Therefore, we define the technological development function, $\sigma_x(t)$, with $t \in \Omega_x$,

$$\sigma_x(t) = \frac{L_x}{1 + e^{-k_x(t - t_{0x})}}, \tag{2}$$

where:

- t_{0x} is when the inflection point is reached (corresponding to the maximum of the first derivative of the function, *i.e.*,

[8] We thank an anonymous referee for pointing out this potential issue.

[9] We use the upload date.

[10] *.optimize.curvefit* is an optimizer. It includes solvers for non-linear problems, linear programming, constrained and non-linear least-squares, root finding, and curve fitting (see, https://docs.scipy.org/doc/scipy/reference/optimize.html).

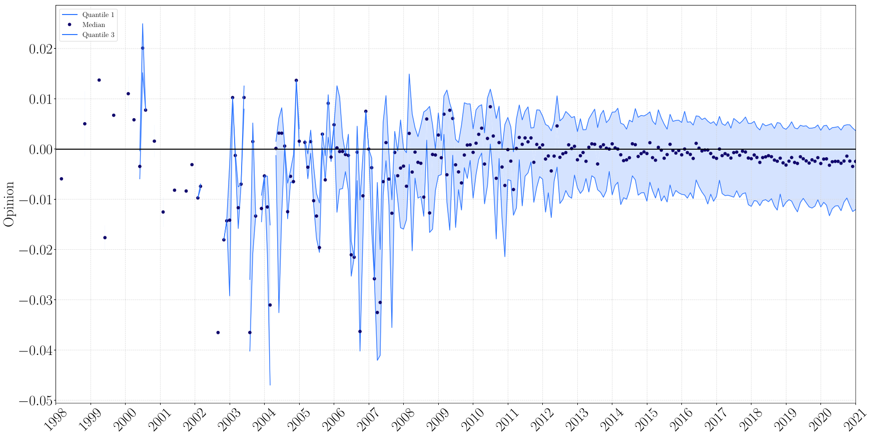

Fig. 5. Distribution of opinion: computer vision and pattern recognition. The median is plotted with dots and the second and third quartiles are plotted with lines. The frequency is monthly and the period is 1998–2021. Similarly to Fig. 4, the plot shows no interesting properties on its left-hand side as the data are sparse. However, a decreasing dispersion (due to the law of large numbers), and a downward trend is present. A multi-plot of all opinion measures (for each CSTCC) available Fig. 12 (Appendix).

the maximum growth rate of technological change (Rogers, 2010));

- L_x is the curve's maximum limit value (*i.e.*, $\lim_{t\to+\infty} \sigma_x(t) = L_x$) (Verhulst, 1838; Rogers, 2010);
- k_x is the sigmoid growth rate or steepness of the curve (Verhulst, 1838; Rogers, 2010).

For every time series of e-prints related to a CSTCC, D_x, the `.optimize.curve_fit` method finds the optimal values of the parameters L_x, k_x and t_{0x} and their standard errors (by minimizing non-linear least-squares errors).[11] If fitting the logistic function to our datasets yields compelling metrics, that is if (i) if 90% of data-points fall within the boundaries of the standard error of the regression, and (ii) if the reduced chi-squared, $\chi^2_{v_x} \approxeq 1$, then we would reject the null hypothesis of no common process (**H1a**).[12]

3.2.2. Security development process

To capture the security attention in each CSTCC, we compute a rolling mean Γ with a window of one year, of the share of e-prints that include at least one word from the security lexicon. We model the rolling mean for each x as follows,

$$\Gamma_{S_{x,t}} = \frac{1}{12} \sum_{i=t-11}^{t} S_{x,i} \tag{3}$$

If the rolling mean displays a positive trend, we would reject the null of (**H1b**). For each CSTCC, we also estimate first order autoregressions to test whether the process is akin to an AR(1) or a random walk with a positive drift.

3.2.3. Relation between technological change and security development

We use a multivariate time-series method for each x concerning the security development and its relationship to technological change. More specifically, we fit a multivariate autoregressive model to each x, which comprises both lagged dependent and independent variables of orders p_x and $q_x \in \Omega_x$). Both orders are determined with the Akaike information criterion (Asteriou and Hall, 2015). We estimate the following specification with OLS,

$$S_{x,t} = \zeta_x + \sum_{i=1}^{p_x} \phi_{x,i} S_{x,t-i} + \sum_{j=1}^{q_x} \theta_{x,j} D_{x,t-j} + u_{x,t}, \tag{4}$$

where:

- S_x is the time series of security development, $S_{x,i}$ is its value at time i, and $\phi_{x,i}$ is its autoregressive parameter;[13]
- D_x is the time series of technological development, $D_{x,j}$ is its value at time j, and $\theta_{x,j}$ is its regressor parameter;
- ζ_x is a constant;
- $u_{x,t}$ is the time series of error terms (*i.e.*, $S_{x,i} - \hat{S}_{x,i}$).

As time series S_x and D_x may include individual seasonal trends for each x, we apply the Seasonal and Trend decomposition using Loess (STL) method (Cleveland et al., 1990). We use a logarithmic transformation on the series as they often present exponential growth curves (Asteriou and Hall, 2015). Finally, as these time series have unit-roots, we differentiate them with order $I_{D_x}(n)$ and $I_{S_x}(m)$, where n and $m \in \mathbb{N}^*$ (Asteriou and Hall, 2015).

If the estimation delivers high adjusted R^2 and statistically significant and positive coefficients across CSTCCs, we would reject the null of no relationship between security development and technological change (**2**).[14]

3.2.4. Opinion mining

We analyze the determinants of the security attention and consider two factors. First, the opinion (**H3a**) and second, the standard deviation of the opinion which measures dispersion (**H3b**). To test **H3a** and **H3b**, we use the cross-sectional approach of Fama & MacBeth (Fama and MacBeth, 1973). This method, originally developed to estimate

[11] Non-linear least squares is the form of least squares analysis used to fit a set of v observations with a model that is non-linear in w unknown parameters ($v \geq w$). The basis of the method is to approximate the model by a linear one and refine the parameters by successive iterations.

[12] In our case, the standard error of the regression is captured by computing the squared root of the reduced chi squared, denoted as χ^2_v. The χ^2_v statistic, also known as the mean squared weighted deviation (MSWD), is used as a goodness-of-fit metric. This statistic can be interpreted as follows: a $\chi^2_v \gg 1$ indicates a poor model fit. A $\chi^2_v > 1$ indicates that the fit has not fully captured the data (or that the error variance has been underestimated). In principle, a value of χ^2_v around 1 indicates that the extent of the match between observations and estimations agrees with the error variance. A $\chi^2_v < 1$ indicates that the model is overfitting the data: either improperly fitting noise or overestimating the error variance (Bevington and Robinson, 2003).

[13] NB: As we capture security development through the security attention, here S_x is used interchangeably for both concepts.

[14] Before interpreting the results, we verify that the time series of $u_{x,t}$ is not (i) serially correlated and (ii) is not heteroskedastic (Asteriou and Hall, 2015). Such statistics are available upon request.

both market-risk exposures and risk premia of assets, is a two-pass estimation. We use the second pass, a sequence of cross-sectional OLS regressions at each month t, with $t \in \Omega_x$, for an x of the form,

$$y_x = \alpha_x + \beta_x \eta_x + \boldsymbol{\gamma_x} \mathbf{Z_x} + \epsilon_x \tag{5}$$

where:

- y_x is the security attention;
- α_x is a constant;
- η_x is the variable of interest (*i.e.*, the opinion (mean or median in turn) and its dispersion, and β_x the coefficient;
- $\mathbf{Z_x}$ is a matrix of additional controls, and γ_x is a vector of coefficients;
- and ϵ_x is the error term.

Next, we consider the estimated time series of β, $\hat{\beta}$, to test whether they significantly depart from zero. In addition, we correct for serial correlation and heteroskedasticity in $\hat{\beta}$ with Newey–West's adjustment method (Newey and West, 1987).

We project the time series of parameters on a constant and extract the covariance matrix of errors that we adjust to retrieve the standard errors. As the procedure of Newey and West (1987) requires knowing the appropriate lag period, we also use the non-parametric approach of Newey and West (1994) with automatic lag selection. Despite the small size of the cross-section (20 CSTCCs), the large time dimension still permits statistical inference.[15] Finally, we consider the unavailability of some CSTCCs at the beginning of our sample and restrict the estimation to a period starting in November 2002, when 15 CSTCCs are simultaneously available in the cross-section (for a total of 217 time-observations in our sub-sample). To proxy for the instantaneous opinion, we consider, in turn, the median and mean computed in each CSTCC. Finally, to control that our results are not driven by the numerator or denominator of the share of security attention, we add two control variables, the number of e-prints with security attention and the total number of e-prints.

4. Results

In this section, we present the results of applying the specified methods (Section 4) we used to test our hypotheses (Section 3). **H1a**, **H1b**, **H3a**, and **H3b** are verified (*i.e.*, null hypotheses are rejected) for all CSTCCs – except for **H1a**, where for 5 out of 20 CSTCCs, the null hypothesis is not rejected –, while **H2** is not verified (*i.e.*, the null hypothesis is not rejected).

4.1. A logistic growth process for technological change

Table 3 shows the metrics and parameter values of non-linear regressions that fit a logistic function to our data.

Out of the 20 CSTCCs, 15 exhibit $\chi^2_\nu > 1$, five exhibit $\chi^2_\nu \gg 1$, and one exhibits $\chi^2_\nu < 1$.[16] Therefore, the logistic function fits our observed data for 15 different CSTCCs. In Fig. 8 we normalize all fits for comparison and we display the logistic fits for all CSTCCs in Fig. 10 (Appendix). Hence, we reject the null of **H1a** for 15 CSTCCs: they follow a common logistic growth process. We interpret the estimates of a $\chi^2_\nu < 1$ and $\chi^2_\nu \gg 1$) as follows. The only estimates $\chi^2_\nu < 1$ is `cs.OS` which has sparse data and thus a high data dispersion (in fact, only 652 e-prints have been uploaded throughout the subcategory's entire history). The estimates $\chi^2_\nu \gg 1$, is obtained for technologies that have not reached their inflection point yet. Thus, in some cases, the Levenberg–Marquardt algorithm does not converge. The CSTCCs concerned are `cs.AI`, `cs.CL`, `cs.LG`, and `cs.SY`. Fig. 6 shows the fit of a typical logistic growth pattern for the CSTCC `cs.DS`, while Fig. 7 shows the fit of a typical exponential growth pattern for the CSTCC `cs.CV`. This exponential growth is typical of the first stage of the logistic growth process. If we cannot confirm that these five CSTCCs will follow a logistic growth process, we cannot reject this hypothesis either. Overall our results support the theory that most CSTCCs follow a common process of technological change (Rogers, 2010).

Table 3
Sigmoid fits of monthly normalized and aggregated number of e-prints per CSTCC.

arXiv categories	χ^2_ν	SE	L	k	t_0
cs.AI	10.062	3.172	49908.902	0.015	2071
cs.AR	1.889	1.375	1297.227	0.016	2065
cs.CC	2.588	1.609	0.489	0.032	2004
cs.CL	12.145	3.485	5.568	0.039	2019
cs.CR	3.013	1.736	10.783	0.015	2028
cs.CV	4.966	2.228	13.626	0.037	2019
cs.DB	2.454	1.567	0.461	0.025	2010
cs.DC	2.178	1.476	1.665	0.020	2015
cs.DS	2.788	1.670	1.273	0.037	2009
cs.GT	1.969	1.403	0.524	0.054	2009
cs.HC	2.275	1.508	1766.652	0.020	2051
cs.IR	2.137	1.462	8.690	0.015	2031
cs.LG	12.952	3.599	12463.478	0.030	2039
cs.NE	3.042	1.744	1.298	0.025	2016
cs.NI	2.383	1.544	1.125	0.046	2008
cs.OS	0.893	0.945	78.830	0.007	2115
cs.PL	2.712	1.647	0.423	0.022	2011
cs.RO	3.762	1.940	6.918	0.032	2022
cs.SE	3.297	1.816	0.855	0.025	2013
cs.SY	10.444	3.232	2.963	0.031	2018

This table displays the goodness-of-fit measures (*i.e.*, the χ^2_ν, and the regression standard error (SE)), and parameters of the sigmoid fits of the total normalized e-prints per CSTCC. The parameter t_0 indicates the year in which the maximum growth rate of the CSTCC is reached.

4.2. Security development increases over time

Fig. 9 shows the rolling mean (Eq. (3)) of the share of security attention.

All CSTCCs display positive trends, depicting an increase in the share of security attention over time. Hence, we reject the null of **H1b**: we find empirical evidence of a security development process which grows for all CSTCCs over time, supporting the view that security attention spikes at a later stage of the technological change.

In Table 4 we report the results of the first-order individual autoregressions. We find a positive coefficient in all but two CSTCCs, thereby confirming the aforementioned graphical results. Moreover, four CSTCCs (cs.AI, cs.CC, cs.CR and cs.NI) are significant at the 1% level, and four others (cs.PL, cs.RO, cs.SE and cs.SY) are significant at the 10% level. We interpret the non-significant positive coefficient of the remaining ones as characteristic of a random walk with positive drift.

4.3. Security development is independent from technological change

We report the results of the tests of **H2** in Table 4. We estimate a multivariate time series regression to test the relation between the security development, $S_{x,t}$, and the technological change, $D_{x,t}$, across CSTCCs, x. We do not find autocorrelation or heteroskedasticity in the error term u for all x (these results are available upon request). We do not find a statistically significant relation between $S_{x,t}$ and $D_{x,t}$: the statistical significance remains well above the 5% threshold for the great majority of regressors of D (*i.e.*, $q **= 0, \forall q, x$). In the case of statistically significant estimates, their magnitude is systematically small (see Table 5). Hence, we cannot reject the null of **H2**: technological change does not explain security development.

[15] In fact, in their study, Fama and MacBeth (1973) use a cross-section of only 20 portfolios.

[16] *i.e.*, the different χ^2_ν lay within the same order of magnitude than 1, and are greater than one for 15 CTSCCs out of 20.

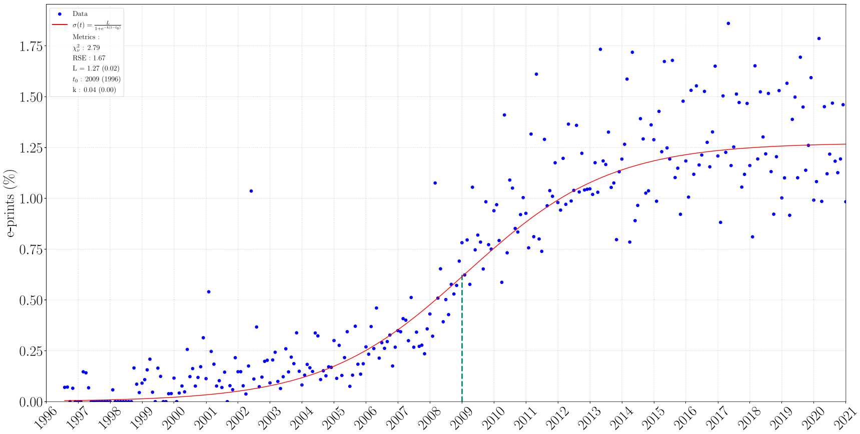

Fig. 6. Logistic growth function fit of e-prints: data structures and algorithms. This figure depicts the normalized e-prints (in blue), and the logistic growth fit (Eq. (2), in red). We additionally plot the inflexion point (vertical green dashed segment). We report the parameters of the fit, their standard errors in parenthesis and the χ^2_ν. The frequency is monthly and the period is 1996–2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

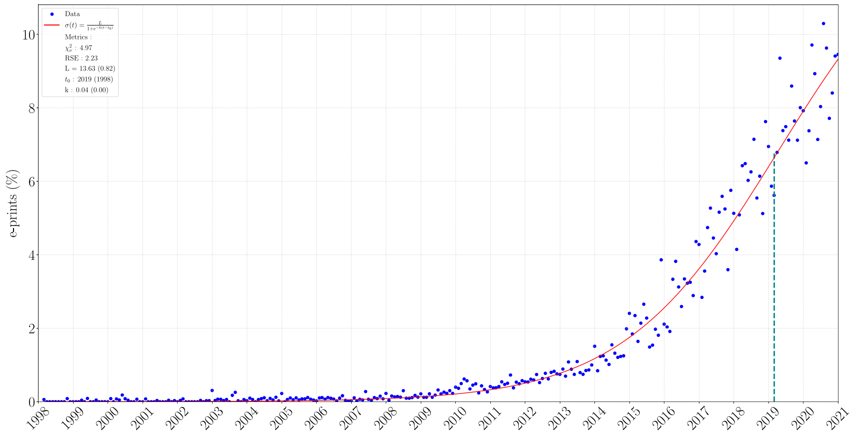

Fig. 7. Beginning of logistic growth of e-prints: computer vision and pattern recognition. This figure depicts the normalized e-prints (in blue), with the logistic growth function fit (Eq. (2), in red). We report the parameters of the fit, the parameters, their standard errors in parenthesis and the χ^2_ν. The frequency is monthly and the period is 1998–2021. In contrast to Fig. 5, the inflection point (*i.e*, t_0) was reached around 2019. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

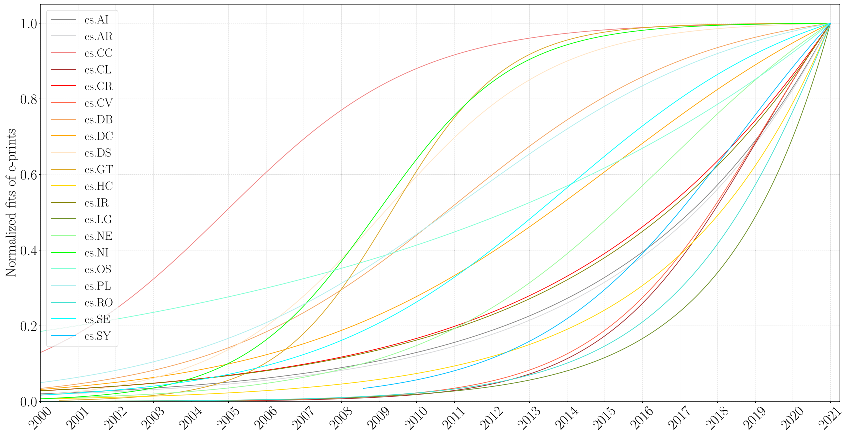

Fig. 8. Normalized fits of logistic growth functions. This figure depicts the fits of all CSTCCs normalized through a division by its maximum. The frequency is monthly and the period is 2000–2021. A 45° line starting from the bottom-left corner to the upper-right, would visually segregate CSTCCs that have reached or not their inflection points above or below the line, respectively. Such a segregation can also be determined by analyzing t_0 of Table 3.

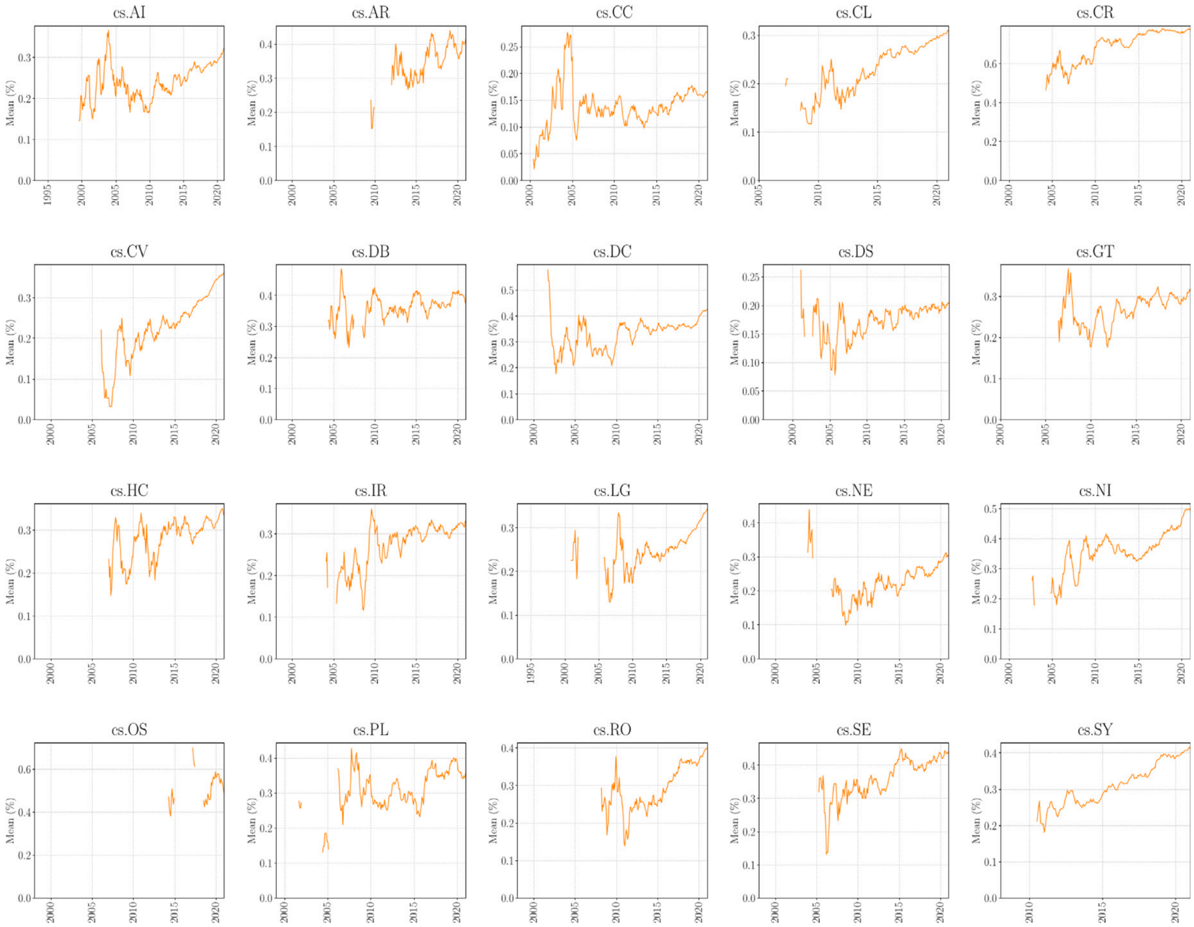

Fig. 9. Multi-plot of the security attention. This figure depicts the share of articles with security attention for each CSTCC. The frequency is monthly, and depending on the CSTCC, the period varies from 1999–2021 and 2014–2021.

Table 4
This table displays the results of individual auto-regressions of order one of the security attention in each of the twenty technologies.

	cs.AI	cs.AR	cs.CC	cs.CL	cs.CR	cs.CV	cs.DB	cs.DC	cs.DS	cs.GT	cs.HC	cs.IR	cs.NE	cs.NI	cs.OS	cs.PL	cs.RO	cs.SE	cs.SY
coefficient	0.30	0.06	0.20	0.07	0.25	0.02	0.07	0.05	0.03	0.02	−0.01	−0.07	0.03	0.34	0.12	0.16	0.13	0.14	0.15
t-statistic	4.40	0.84	3.08	1.01	3.89	0.26	1.04	0.73	0.49	0.25	−0.18	−1.00	0.45	5.31	1.41	2.31	1.78	2.07	1.77

4.4. Opinion, consensus, and security development

We report the results of the tests of **H3a** and **H3b** in Table 6. We test the relation between the instantaneous and aggregate measure of opinion (mean and median in turn) and the share of security attention across CSTCCs (**H3a**) with a Fama–Macbeth approach. In all specifications, we document a significant negative relation that holds after the Newey–West adjustment (bandwidth ranging between 1 and 5, and truncated for the lag selection). More specifically, in the parsimonious version of the model, the estimates of security attention is significant at the 1% level (t-stats of 4.00 and 3.37 for the mean and the median, respectively). These results are robust to the inclusion of the (log) number of e-prints and (log) number of e-prints containing security attention. The size of the estimates remains close and the statistical significance remains well below the 1% threshold. Thus, we rule out the possibility that our results are driven by either the numerator or the denominator used to construct the variable of interest. Moreover, given that we employ a cross-sectional methodology, these specifications also discard the possibility of a spurious time-effect as an explanation for our results. Interestingly, the point estimate for the numerator is positive and significant at the 1% level in all specifications, while that of the denominator is highly significant in the last specification only. This makes us confident that the share of security attention variable is different from the absolute number of e-prints with security attention and total number of e-prints. Hence, we reject the null of (**H3a**). The experts' opinion in a given CSTCC is significantly and negatively related to the security attention.

In a final specification, we include the standard deviation of the opinion as an explanatory variable to proxy for the (inverse) consensus. We obtain similar orders of magnitude for the estimate and a statistical significance that remains on par with the usual significance thresholds. The estimates are highly significant and positive (adjusted t-statistic up to 7.67). These results align with those of the literature in finance. Such a positive relation between opinion dispersion (of *e.g.*, analysts who provide price targets and recommendations for stocks) and actual stock returns is well documented (Diether et al., 2002) and theoretically backed by asset pricing models (Johnson, 2004). Hence, we reject the null of (**H3b**): The consensus (dispersion of opinion) is negatively (positively) related to the security attention. Consequently, we find empirical evidence for two contemporaneous determinants of the opinion towards the CSTCCs.

5. Discussion

In this section, we discuss the social implications and the academic relevance of our findings, as well as paths for further research.

Table 5
Multivariate time-series regression of security development.

		$S_{x,t}$				$D_{x,t}$			Adjusted R^2	SE regression	AIC	Sum resid^2	F-stat
		p_x	p_x^{**}	I_{S_x}	interpolated ratio	q_x	q_x^{**}	I_{D_x}					
$S_{x,t}$	cs.AI	12	7	1	0.04	1	0	2	0.419	0.099	−556.9	2.98	18.63
	cs.AR	12	8	1	0.23	2	2	1	0.448	0.139	−272.0	4.69	15.88
	cs.CC	12	10	1	0.24	1	0	1	0.478	0.060	−976.5	1.24	25.96
	cs.CL	12	8	1	0.02	1	0	2	0.436	0.094	−571.2	2.65	19.46
	cs.CR	12	5	1	0.12	1	0	1	0.508	0.095	−524.7	2.47	23.82
	cs.CV	12	8	1	0.15	1	0	3	0.569	0.111	−401.6	3.07	27.79
	cs.DB	12	7	1	0.07	1	1	1	0.480	0.113	−378.8	3.12	19.26
	cs.DC	12	8	1	0.03	1	0	1	0.508	0.104	−421.9	2.64	21.41
	cs.DS	12	8	1	0.22	1	0	1	0.419	0.075	−783.8	1.81	19.78
	cs.GT	12	8	1	0.11	1	0	1	0.403	0.107	−374.6	2.54	13.28
	cs.HC	12	10	2	0.14	1	0	2	0.739	0.140	−266.8	4.78	56.87
	cs.IR	11	6	1	0.06	1	0	2	0.624	0.115	−373.7	3.24	36.74
	cs.LG	12	8	1	0.09	1	0	2	0.420	0.105	−448.3	2.88	16.32
	cs.NE	12	7	1	0.11	2	0	1	0.549	0.111	−402.1	3.07	24.02
	cs.NI	12	9	1	0.09	7	4	1	0.398	0.111	−411.3	3.15	10.55
	cs.OS	12	4	1	0.35	2	1	1	0.431	0.147	−241.4	5.28	14.93
	cs.PL	12	8	1	0.20	2	0	1	0.437	0.099	−544.8	2.95	18.39
	cs.RO	12	5	1	0.23	1	0	1	0.350	0.104	−420.8	2.60	11.54
	cs.SE	12	8	1	0.11	2	0	1	0.475	0.105	−418.6	2.65	17.60
	cs.SY	12	4	1	0.26	1	0	2	0.379	0.079	−420.1	1.11	10.02

This table lists the respective (i) autoregression order, p_x, and regression order, q_x, (ii) number of statistically significant (up to $p > 0.05$) autoregressors, p_x^{**}, and regressors, q_x^{**}, and (iii) the degree of differentiation, I_{S_x} and I_{D_x}. We also report the interpolated ratio of S, as these time series cannot present null values (otherwise, the share of *security considerations* would be zero). To fulfill missing values, we perform a linear interpolation between concerned data points. Regression metrics are on the right.

Table 6
Cross-sectional regressions average of security attention.

	Security attention								
Mean opinion	−4.56 (−6.73) [−4.79]	−3.68 (−5.70) [−3.52]	−4.73 (−7.03) [−4.71]				−3.00 (−4.25) [−2.86]	−2.04 (−3.07) [−1.80]	−1.83 (−6.63) [−4.49]
Opinion σ				9.84 (12.30) [7.67]	6.10 (7.26) [3.71]	10.41 (12.74) [7.37]	9.40 (11.35) [7.41]	6.07 (6.96) [3.92]	0.92 (2.76) [2.64]
Log (# e-prints with security attention)		0.13 (15.22) [5.87]			0.13 (16.27) [6.49]			0.13 (15.11) [6.27]	
Log (# e-prints)			−0.004 (−0.71) [−0.63]			−0.01 (−1.39) [−1.58]			0.38 (70.83) [55.53]
Average R^2	0.13	0.39	0.22	0.15	0.42	0.22	0.25	0.50	0.92

	Security attention								
Median opinion	−4.23 (−6.23) [−4.25]	−3.49 (−5.54) [−3.34]	−4.49 (−6.63) [−4.56]				−3.31 (−4.89) [−3.15]	−2.24 (−3.52) [−1.97]	−2.17 (−7.91) [−5.42]
Opinion σ				9.84 (12.30) [7.67]	6.10 (7.26) [3.71]	10.41 (12.74) [7.37]	10.01 (12.19) [7.63]	6.34 (7.01) [3.90]	1.01 (3.04) [2.86]
Log (# e-prints with security attention)		0.13 (15.43) [5.91]			0.13 (16.27) [6.49]			0.13 (15.45) [6.31]	
Log (# Total e-prints)			−0.004 (−0.58) [−0.49]			−0.01 (−1.39) [−1.58]			0.38 (70.52) [55.21]
Average R^2	0.12	0.38	0.21	0.15	0.42	0.22	0.25	0.49	0.92

This Table reports the time-series average of parameters from cross-sectional regressions of security attention. The explanatory variables are the mean opinion, standard deviation of opinion, (log) number of e-prints with security attention, and (log) number of e-prints. The first (second) panel presents the results with mean (median) opinion. We report unadjusted t-statistics in parenthesis, and Newey–West (1994) t-statistics in square brackets. We additionally report the average R^2, for each specification. The sample period is January 2002–November 2020 and the number of time-series observations is 224.

5.1. Social implications

Whether it ensures the privacy of health data, the robustness and continuity of governments and critical infrastructures, or the network stability in times of war, digital trust has become a central concern in highly digitized societies. Yet, as emphasized by the World Economic Forum, building sustainable digital trust requires the development of robust metrics. This process starts at the inception of an information technology. We use a combination of logistic growth modeling, lexicon-based topic extraction, and opinion mining to measure how digital trust changes throughout the life-cycle of a technology. We find that the experts' opinion explains best the progressive buildup of digital trust of information technologies, although the increase of digital trust is independent of technological change.

Our results are important as they represent a first quantitative measure of the evolution of information technology and digital trust.

The practical use of our quantitative framework includes technological roadmapping for strategic management of information systems, as well as policy making. Our proposed methods should help pilot the acquisition and deployment of information technologies as well as help manage research efforts to further build trust in a specific technology. For instance, computer vision has numerous applications in both civil and military fields. Yet, for this specific technology, we find that digital trust currently decreases, while the technology development has just passed the logistic inflexion point. As they pervade society, these technologies will require close digital trust scrutiny.

5.2. *Academic importance*

First, the logistic growth process of technological change identified for the great majority of the 20 CSTCCs enables us to compare them through a common process, and to forecast their future growth trajectories. Knowing when a CSTCC reaches its inflection point helps to anticipate a CSTCC's slow-down and maturity (Rogers, 2010), and even its obsolescence (Parvin, 2017). This indicator helps to prioritize investment and acquisition of technologies, an opportunity-cost challenge. Knowing technologies' growth stages helps to time investment decisions appropriately (Keupp et al., 2019). Information about the upcoming slow down of CSTCC's growth, relative to other technologies or in absolute terms, is critical to set priorities. For instance, organizations willing to invest in emerging (mature) technologies are likely to prioritize technologies that have not reached (reached) their inflection point.

Second, the lack of a significant relationship between technological change and security development, and the fact that the latter occurs at a late stage, supports the view of a lack of security standards in computer science (Casola et al., 2020; Kreitz, 2019; Böhme, 2013; Anderson, 2001; Panarello et al., 2018). Our approach also helps to evaluate the security development among different technologies (Anderson, 2020). Decision-makers who acquire technologies must grasp the security maturity level, in particular given that security is often discarded at the expense of revenue or user-base growth (Anderson, 2020). How much attention has been paid to security development during technological change stages? When did the security development pick up? Our approach sheds light on these aspects and may help IT decision-makers, such as CISOs, to prioritize investments or acquisitions. We also give hints on how aspects related to security are considered within the engineering process.

Third, to complement the investigation of the technological change and security development, we extract experts' opinion towards technologies. We investigate how opinion and consensus affects security development. This relation improves decision-making in technologies' selection. Similar to financial assets, for which opinion explains returns, the opinion associated with technologies are linked to security risks (Diether et al., 2002). The relation between the dispersion of opinion and the security attention is another striking example of how security risks can be explained with opinion measures.

5.3. *Future research*

We use the measure of scientific works as a proxy for technological change in alignment with an abundant bibliometric literature (Dotsika and Watkins, 2017; Jaewoo and Woonsun, 2014; Rezaeian et al., 2017). More specifically, we measure the number of e-prints uploaded through time in the `arXiv` repository for each CSTCC. With this measure, we capture the scientific community's attention to CSTCCs. In the field of computer science, it is a common practice to upload e-prints on the `arXiv` repository whenever they are ready for submission to a scientific venue. The `arXiv` repository is thus considered by computer scientists as a central repository. Consequently, we argue that the latest scientific advances are captured by the `arXiv` platform. However, other providers such as `OpenAlex`, `Dimensions AI`, `Scopus`, `Web of Science`, or `Semantic Scholar` might be considered as well.[17] Also, other aspects and measures of technological development may be used, such as technology adoption (*e.g.*, by considering the growth in software/hardware instances and the number of users, and/or by considering the number of patents and the dynamics of social-media heuristic recurrences), or technology maturity (*e.g.*, by considering the opinion of users in their reviews, or TRL measurements).

Concerning the share of security attention expressed in e-prints, we select a set of keywords related to concepts depicted in the *Information Security* portal of *Wikipedia*. These concepts relate to the CIA triad (*i.e.*, *confidentiality*, *integrity*, and *availability*) and the *non-repudiation* principle (Cherdantseva and Hilton, 2013; Ritzdorf et al., 2017). Our results support the hypothesis that the security attention improves at a late stage of the technological change process. We leave for future research the investigation and measure of how substantial this delay between technological development and security development is. This could be done by implementing a delay function, which could also be employed to investigate whether a *catch-up effect* is present and for which CSTCCs this effect takes place. However, one might argue that such a pattern in *security attention* is induced by an omitted variable: the overall growing interest in cybersecurity issues. Unfortunately, accounting for such an omitted variable seems hardly feasible in practice. Yet, when we conducted the Fama–MacBeth (cross-sectional analysis), which is not influenced by any time trend – and thus by such an omitted variable –, we found support for the relation between security attention and opinion. This demonstrates that this omitted variable does not affect the test of **H3a**. Future research could enhance the selection of keywords related to security attention by implementing other information retrieval methods such as *tf-idf* or *Key-BERT* to capture the recurrence of words related to security, and use the most recurrent ones as filters to capture the security attention in e-prints.

Finally, our classic lexicon-based approach to capture opinion could be enhanced with machine-learning approaches (*e.g.*, decision-trees, support-vector machines, or neural networks). Such approaches would yield a higher precision for sentiment analysis. However, researchers willing to use such sophisticated NLP methods may face issues finding labeled datasets. For instance, *BERT* (Tenney et al., 2019) and *XLNet* (Myagmar et al., 2019) models, despite being pre-trained with a plethora of datasets, are not directly transferable to datasets presenting other text structures. Unfortunately, to the best of our knowledge, there are no academic-work datasets labeled for sentiment analysis.

6. Conclusion

Little work has been done to model a holistic and dynamic indicator that captures the overall change in a technology's security level — especially with respect to technological change. We conceptualize and measure such an indicator based on the investigation of what we call security development. This indicator is described by (i) the statistical relation between technological change and security development, (ii) the security development itself, modeled as the evolution of security attention of different technologies, and (iii) the effect of opinion and consensus on security development. We adopt a bibliometric approach related to 20 computer-science technology categories. Our results bring a unique view on the technological change and security development processes. First, we are not able to find a dependence between the two processes. Second, we find that more attention is paid to security at a late stage of technological change. Third, we find that the opinion is a significant and negative determinant of the security attention. Our indicator also evaluates the efficiency of the NIST recommendation to shift towards open security, i.e. the use of open-source philosophies and methodologies to approach security information technologies. To our

[17] https://docs.openalex.org/api

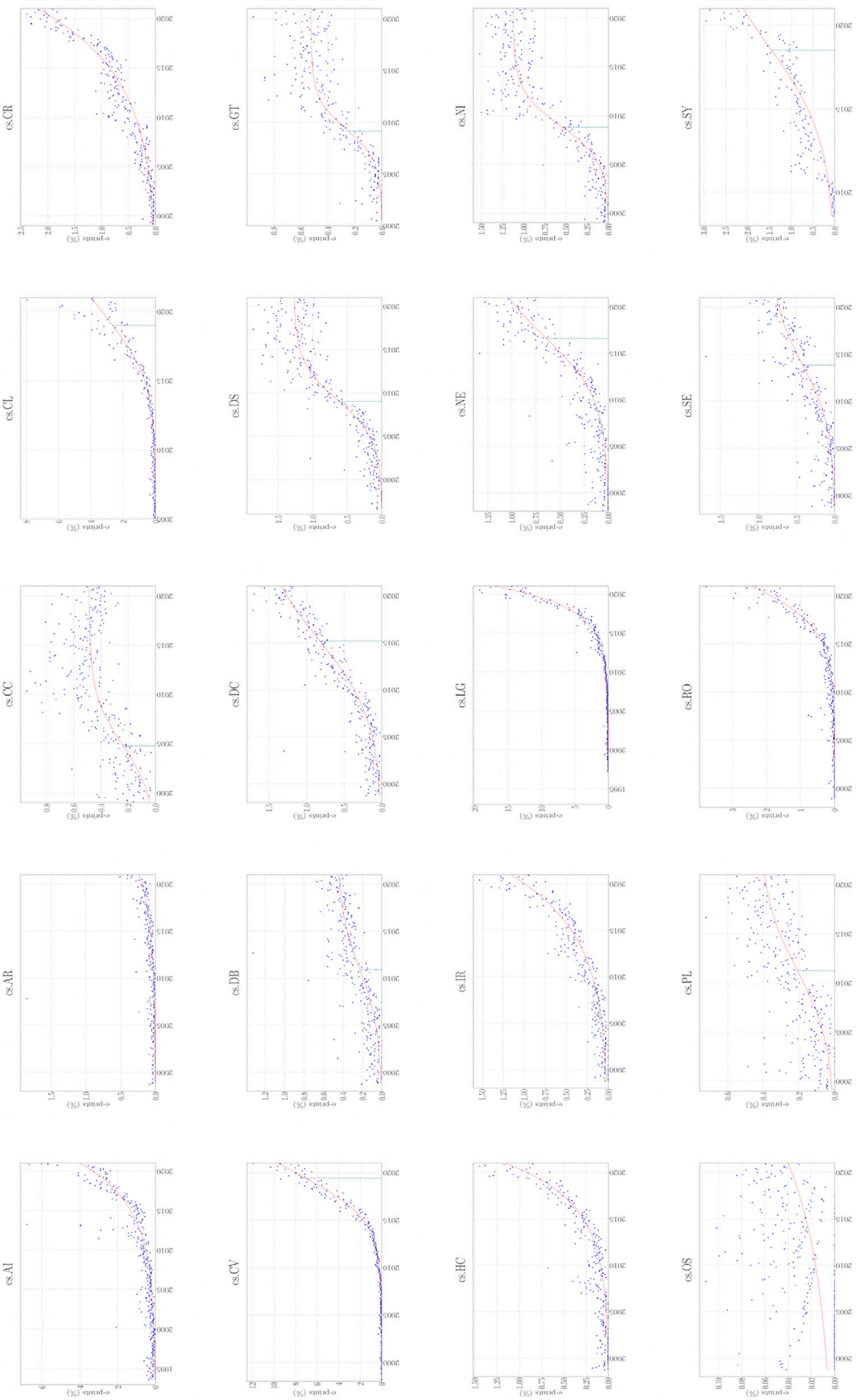

Fig. 10. Multi-plot of logistic growth fits.

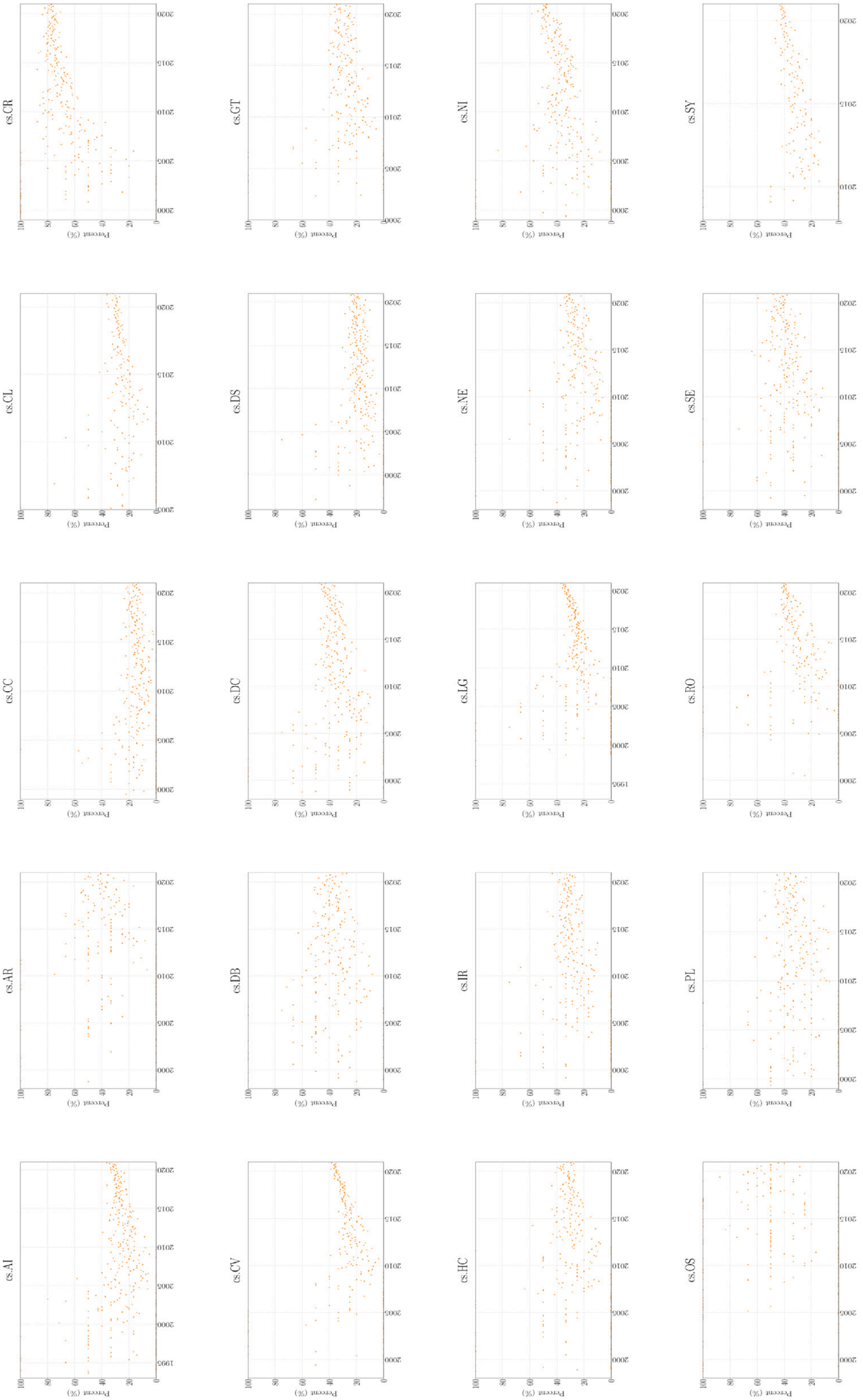

Fig. 11. Multi-plot of security attention.

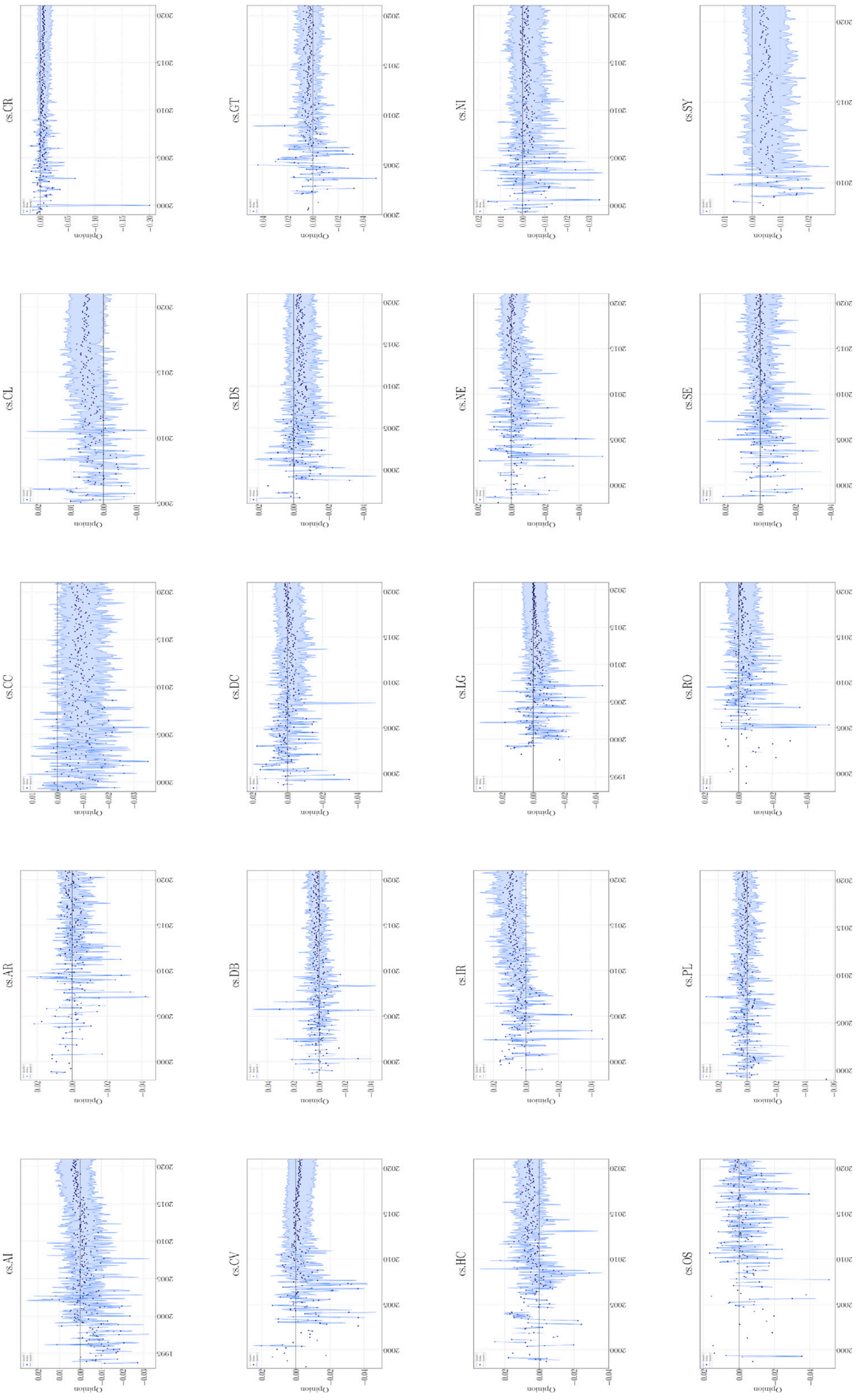

Fig. 12. Multi-plot of opinion.

knowledge, our work presents for the first time an indicator capable of capturing open-security dynamics in computer-science technologies. We are thus able to show for instance that even in an open-security context, security-by-design is not common practice. Furthermore, our benchmark may measure the progress towards this open-security recommendation over time. We leave for future research to measure if this recommendation accelerates a *security-by-design* approach given a technology at a low technology-readiness level. Altogether, this research brings new methods to model the security development of computer-science technologies.

CRediT authorship contribution statement

Dimitri Percia David: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Loïc Maréchal:** Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. **William Lacube:** Data curation, Formal analysis, Investigation, Software, Visualization. **Sébastien Gillard:** Data curation, Investigation, Software, Visualization. **Michael Tsesmelis:** Writing – review & editing. **Thomas Maillart:** Conceptualization, Supervision, Writing – review & editing. **Alain Mermoud:** Conceptualization, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This research received a Cyber-Defence (CYD) Campus Fellowship, through the EPFL Research Office.

Data availability

Please find the link for the code and data here: https://github.com/technometrics-lab/1-Security_Dynamics

arXiv dataset (Original data) (GitHub)

Appendix

See Figs. 10–12.

References

Abernathy, W.J., Utterback, J.M., et al., 1978. Patterns of industrial innovation. Technol. Rev. 80 (7), 40–47.

Adamuthe, A.C., Tomke, J.V., Thampi, G.T., 2014. Technology forecasting: The case of cloud computing and sub-technologies. Int. J. Comput. Appl. 106 (2), 14–19.

Adner, R., Levinthal, D.A., 2002. The emergence of emerging technologies. Calif. Manage. Rev. 45 (1), 50–66.

Akerlof, G.A., 1978. The market for "lemons": Quality uncertainty and the market mechanism. In: Uncertainty in Economics. Elsevier, pp. 235–251.

Ali, Z., Qi, G., Kefalas, P., Khusro, S., Khan, I., Muhammad, K., 2022. SPR-SMN: Scientific paper recommendation employing SPECTER with memory network. Scientometrics 1–23.

An, J., Kim, K., Mortara, L., Lee, S., 2018. Deriving technology intelligence from patents: Preposition-based semantic analysis. J. Informetr. 12, 217–236.

An, X., Sun, X., Xu, S., 2022. Important citations identification with semi-supervised classification model. Scientometrics 1–23.

Andersen, B., 1999. The hunt for S-shaped growth paths in technological innovation: A patent study. J. Evol. Econ. 9 (4), 487–526. http://dx.doi.org/10.1007/s001910050093.

Anderson, R., 2001. Why information security is hard - an economic perspective. In: Seventeenth Annual Computer Security Applications Conference. IEEE, pp. 358–365.

Anderson, R., 2020. Security Engineering: A Guide to Building Dependable Distributed Systems, third ed. Wiley.

Anderson, R., Moore, T., 2006. The economics of information security. Science 314 (5799), 610–613.

Anderson, R., Moore, T., 2007. The economics of information security: A survey and open questions. In: Fourth Bi-Annual Conference on the Economics of the Software and Internet Industries. Toulouse School of Economics, pp. 19–20.

Antons, D., Grünwald, E., Cichy, P., Salge, T.O., 2020. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. R D Manag. 50, 329–351.

Assante, M.J., Tobey, D.H., 2011. Enhancing the cybersecurity workforce. IT Prof. 13 (1), 12–15.

Asteriou, D., Hall, S.G., 2015. Applied Econometrics. Macmillan International Higher Education.

Bai, X., 2011. Predicting consumer sentiments from online text. Decis. Support Syst. 50 (4), 732–742. http://dx.doi.org/10.1016/j.dss.2010.08.024.

Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technol. Forecast. Soc. Change 73 (7), 835–844. http://dx.doi.org/10.1016/j.techfore.2005.09.001.

Bevington, P.R., Robinson, D.K., 2003. Data Reduction and Error Analysis. McGraw Hill.

Böhme, R., 2013. The Economics of Information Security and Privacy. Springer-Verlag, http://dx.doi.org/10.1007/978-3-642-39498-0.

Brock, G.W., 2021. The Second Information Revolution. Harvard University Press.

Calleja-Sanz, G., Olivella-Nadal, J., Solé-Parellada, F., 2020. Technology forecasting: Recent trends and new methods. Res. Methodol. Manag. Ind. Eng. 45–69.

Carlton, M., 2016. Development of a cybersecurity skills index: a scenarios-based, hands-on measure of non-IT professionals' cybersecurity skills (Ph.D. thesis). NSUWorks.

Casola, V., De Benedictis, A., Rak, M., Villano, U., 2020. A novel security-by-design methodology: Modeling and assessing security by SLAs with a quantitative approach. J. Syst. Softw. 163, 1–56.

Chang, W.-L., Wang, J.-Y., 2018. Mine is yours? Using sentiment analysis to explore the degree of risk in the sharing economy. Electron. Commer. Res. Appl. 28, 141–158. http://dx.doi.org/10.1016/j.elerap.2018.01.014.

Chen, Y.-H., Chen, C.-Y., Lee, S.-C., 2011. Technology forecasting and patent strategy of hydrogen energy and fuel cell technologies. Int. J. Hydrogen Energy 36 (12), 6957–6969. http://dx.doi.org/10.1016/j.ijhydene.2011.03.063.

Chen, H., Song, X., Jin, Q., Wang, X., 2022. Network dynamics in university-industry collaboration: A collaboration-knowledge dual-layer network perspective. Scientometrics 1–24.

Chen, H., Zhang, G., Zhu, D., Lu, J., 2017. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. Technol. Forecast. Soc. Change 119, 39–52. http://dx.doi.org/10.1016/j.techfore.2017.03.009.

Cherdantseva, Y., Hilton, J., 2013. A reference model of information assurance security. In: 2013 International Conference on Availability, Reliability and Security. pp. 546–555. http://dx.doi.org/10.1109/ARES.2013.72.

Choi, J., Hwang, Y.-S., 2014. Patent keyword network analysis for improving technology development efficiency. Technol. Forecast. Soc. Change 83, 170–182. http://dx.doi.org/10.1016/j.techfore.2013.07.004.

Choi, J., Lee, J., Yoon, J., Jang, S., Kim, J., Choi, S., 2022. A two-stage deep learning-based system for patent citation recommendation. Scientometrics 1–22.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: A seasonal-trend decomposition. J. Off. Stat. 6 (1), 3–73.

Coccia, M., 2005. Technometrics: Origins, historical evolution and new directions. Technol. Forecast. Soc. Change 72 (8), 944–979. http://dx.doi.org/10.1016/j.techfore.2005.05.011.

Daim, T.U., Chiavetta, D., Porter, A.L., Saritas, O., 2016. Anticipating future innovation pathways through large data analysis. Springer.

Daim, T., Iskin, I., Li, X., Zielsdorff, C., Bayraktaroglu, A.E., Dereli, T., Durmusoglu, A., 2012. Patent analysis of wind energy technology using the patent alert system. World Pat. Inf. 34 (1), 37–47. http://dx.doi.org/10.1016/j.wpi.2011.11.001.

Daim, T., Yalçin, H., 2022. Digital Transformations: New Tools and Methods for Mining Technological Intelligence. Edward Elgar Publishing.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. J. Finance 57 (5), 2113–2141.

Dotsika, F., Watkins, A., 2017. Identifying potentially disruptive trends by means of keyword network analysis. Technol. Forecast. Soc. Change 119, 114–127.

Dou, R., Zhang, Y., Nan, G., 2017. Iterative product design through group opinion evolution. Int. J. Prod. Res. 55 (13), 3886–3905. http://dx.doi.org/10.1080/00207543.2017.1316020.

Erzurumlu, S.S., Pachamanova, D.A., 2020. Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. Technol. Forecast. Soc. Change 156, 120041.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. J. Polit. Econ. 81 (3), 607–636.

Fang, X., Zhan, J., 2015. Sentiment analysis using product review data. J. Big Data 2 (1), 1–14. http://dx.doi.org/10.1186/s40537-015-0015-2.

Feng, L., Wang, Q., Wang, J., Lin, K.-Y., 2022. A review of technological forecasting from the perspective of complex systems. Entropy 24 (6), 787.

Golembiewski, B., vom Stein, N., Sick, N., Wiemhöfer, H.-D., 2015. Identifying trends in battery technologies with regard to electric mobility: Evidence from patenting activities along and across the battery value chain. J. Clean. Prod. 87, 800–810. http://dx.doi.org/10.1016/j.jclepro.2014.10.034.

Goode, J., Levy, Y., Hovav, A., Smith, J., 2018. Expert assessment of organizational cybersecurity programs and development of vignettes to measure cybersecurity countermeasures awareness. Online J. Appl. Knowl. Manag. (OJAKM) 6 (1), 54–66.

Guo, W., Wang, H., Tian, Y., Xian, M., 2019. Research on cyberspace security testing and evaluation technology development trend. In: 2019 International Conference on Communications, Information System and Computer Engineering. CISCE, pp. 363–367.

Gurung, A., Raja, M., 2016. Online privacy and security concerns of consumers. Inf. Comput. Security 24 (4), 348–371. http://dx.doi.org/10.1108/ICS-05-2015-0020.

Hajikhani, A., Suominen, A., 2022. Mapping the sustainable development goals (SDGs) in science, technology and innovation: Application of machine learning in SDG-oriented artefact detection. Scientometrics 1–33.

Haleem, A., Mannan, B., Luthra, S., Kumar, S., Khurana, S., 2019. Technology forecasting (TF) and technology assessment (TA) methodologies: A conceptual review. Benchmarking Int. J. 26 (1), 48–72. http://dx.doi.org/10.1108/BIJ-04-2018-0090.

Han, J., Moraga, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: From Natural to Artificial Neural Computation. In: Lecture Notes in Computer Science, Springer, pp. 195–201. http://dx.doi.org/10.1007/3-540-59497-3_175.

Hao, J., Yan, Y., Gong, L., Wang, G., Lin, J., 2014. Knowledge map-based method for domain knowledge browsing. Decis. Support Syst. 61, 106–114. http://dx.doi.org/10.1016/j.dss.2014.02.001.

Hong, S., Kim, J., Woo, H.-G., Kim, Y.-C., Lee, C., 2021. Screening ideas in the early stages of technology development: A word2vec and convolutional neural network approach. Technovation 112, 102407.

Howard, M., Lipner, S., 2006. The Security Development Lifecycle, eightth ed. Microsoft Press Redmond.

Huang, J., Boh, W.F., Goh, K.H., 2019. Opinion convergence versus polarization: Examining opinion distributions in online word-of-mouth. J. Assoc. Inf. Sci. Technol. 70 (11), 1183–1193. http://dx.doi.org/10.1002/asi.24193, _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24193.

Hubbard, D.W., Seiersen, R., 2016. How to Measure Anything in Cybersecurity Risk. John Wiley & Sons.

Jaewoo, C., Woonsun, K., 2014. Themes and trends in Korean educational technology research: A social network analysis of keywords. Procedia - Soc. Behav. Sci. 131, 171–176. http://dx.doi.org/10.1016/j.sbspro.2014.04.099.

Jaffe, A.B., Newell, R.G., Stavins, R.N., 2002. Environmental policy and technological change. Environ. Res. Econ. 22 (1), 41–70.

Jang-Jaccard, J., Nepal, S., 2014. A survey of emerging threats in cybersecurity. J. Comput. System Sci. 80 (5), 973–993.

Johnson, T.C., 2004. Forecast dispersion and the cross section of expected returns. J. Finance 59 (5), 1957–1978. http://dx.doi.org/10.1111/j.1540-6261.2004.00688.x.

Jun, S., Sung Park, S., Sik Jang, D., 2012. Technology forecasting using matrix map and patent clustering. Ind. Manag. Data Syst. 112 (5), 786–807. http://dx.doi.org/10.1108/02635571211232352.

Keupp, M.M., Percia David, D., Mermoud, A., 2019. Militärökonomie. Springer.

Klepper, S., 1997. Industry life cycles. Ind. Corp. Chang. 6 (1), 145–182.

Kreitz, M., 2019. Security by design in software engineering. ACM SIGSOFT Softw. Eng. Notes 44 (3), 23.

Laube, S., Böhme, R., 2017. Strategic aspects of cyber risk information sharing. ACM Comput. Surv. 50 (5), 1–36.

Lee, C., 2021. A review of data analytics in technological forecasting. Technol. Forecast. Soc. Change 166, 120646. http://dx.doi.org/10.1016/j.techfore.2021.120646.

Lee, C., Lee, G., 2019. Technology opportunity analysis based on recombinant search: Patent landscape analysis for idea generation. Scientometrics 121, 603–632.

Lee, P.-C., Su, H.-N., Wu, F.-S., 2010. Quantitative mapping of patented technology: The case of electrical conducting polymer nanocomposite. Technol. Forecast. Soc. Change 77 (3), 466–478. http://dx.doi.org/10.1016/j.techfore.2009.08.006.

Lehrer, K., Wagner, C., 2012. Rational Consensus in Science and Society: A Philosophical and Mathematical Study, twentyfourth ed. Springer Science & Business Media.

Li, L., He, W., Xu, L., Ash, I., Anwar, M., Yuan, X., 2019a. Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. Int. J. Inf. Manage. 45, 13–24.

Li, X., Xie, Q., Daim, T., Huang, L., 2019b. Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology. Technol. Forecast. Soc. Change 146, 432–449. http://dx.doi.org/10.1016/j.techfore.2019.01.012.

Liu, B., 2012. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. 1–167.

Liu, Y., Sarabi, A., Zhang, J., Naghizadeh, P., Karir, M., Bailey, M., Liu, M., 2015. Cloudy with a chance of breach: forecasting cyber security incidents. In: 24th USENIX Security Symposium (USENIX Security 15). USENIX, pp. 1009–1024.

Lotfi, A., Lotfi, A., Halal, W.E., 2014. Forecasting technology diffusion: A new generalisation of the logistic model. Technol. Anal. Strateg. Manag. 26 (8), 943–957. http://dx.doi.org/10.1080/09537325.2014.925105.

Maks, I., Vossen, P., 2013. Sentiment analysis of reviews: Should we analyze writer intentions or reader perceptions? In: Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013. INCOMA, pp. 415–419.

Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., Mutschke, P., 2014. Bibliometric-enhanced information retrieval. In: European Conference on Information Retrieval. Springer, pp. 798–801.

Meland, P.H., Tokas, S., Erdogan, G., Bernsmed, K., Omerovic, A., 2021. A systematic mapping study on cyber security indicator data. Electronics 10 (9), 1092–1118.

Mikheev, A.V., 2020. Technological forecasting related to the energy sector: A scientometric overview. E3S Web Conf. 209, 1–5. http://dx.doi.org/10.1051/e3sconf/202020902022.

Moré, J.J., 1978. The Levenberg-Marquardt algorithm: Implementation and theory. In: Numerical Analysis. Springer, pp. 105–116.

Myagmar, B., Li, J., Kimura, S., 2019. Cross-domain sentiment classification with bidirectional contextualized transformer language models. IEEE Access 7, 163219–163230.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708. http://dx.doi.org/10.2307/1913610.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econom. Stud. 61 (4), 631–653. http://dx.doi.org/10.2307/2297912.

Noh, H., Song, Y.-K., Lee, S., 2016. Identifying emerging core technologies for the future: Case study of patents published by leading telecommunication organizations. Telecommun. Policy 40 (10), 956–970. http://dx.doi.org/10.1016/j.telpol.2016.04.003.

Panarello, A., Tapas, N., Merlino, G., Longo, F., Puliafito, A., 2018. Blockchain and IOT integration: A systematic survey. Sensors 18 (8), 2575–2612.

Parvin, Jr., A.J., 2017. Forecasting technology obsolescence: assessing the existing literature, a systematic review. In: Proceedings of the International Annual Conference of the American Society for Engineering Management. American Society for Engineering Management (ASEM), pp. 1–13.

Perez, C., 2010. Technological revolutions and techno-economic paradigms. Camb. J. Econ. 34 (1), 185–202.

Perez, C., et al., 2010. The Financial Crisis and the Future of Innovation: a View of Technical Change with the Aid of History. Technical Report, TUT Ragnar Nurkse Department of Innovation and Governance.

Pletea, D., Vasilescu, B., Serebrenik, A., 2014. Security and emotion: Sentiment analysis of security discussions on github. In: Proceedings of the 11th Working Conference on Mining Software Repositories. Association for Computing Machinery, pp. 348–351.

Porter, A.L., Roper, A.T., Mason, T.W., Rossini, F.A., Banks, J., 2011. Forecasting and management of technology. John Wiley & Sons.

Priestley, M., Sluckin, T.J., Tiropanis, T., 2020. Innovation on the web: The end of the S-curve? Internet Hist. 4 (4), 390–412.

Rezaeian, M., Montazeri, H., Loonen, R.C.G.M., 2017. Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation. Technol. Forecast. Soc. Change 118, 270–280. http://dx.doi.org/10.1016/j.techfore.2017.02.027.

Ritzdorf, H., Wüst, K., Gervais, A., Felley, G., Capkun, S., 2017. TLS-N: Non-repudiation over TLS enabling-ubiquitous content signing for disintermediation. Cryptol. EPrint Arch. 2017 (578), 1–16.

Rogers, E.M., 1995. Diffusion of innovations: Modifications of a model for telecommunications. In: Schriftenreihe Des Wissenschaftlichen Instituts Für Kommunikationsdienste, Vol. 17. Springer, pp. 25–38.

Rogers, E.M., 2010. Diffusion of Innovations. Simon and Schuster.

Santos, J.C.S., Tarrit, K., Mirakhorli, M., 2017. A catalog of security architecture weaknesses. In: 2017 IEEE International Conference on Software Architecture Workshops. ICSAW, pp. 220–223. http://dx.doi.org/10.1109/ICSAW.2017.25.

Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E., 2015. Sentiment analysis: A review and comparative analysis of web services. Inform. Sci. 311, 18–38.

Shalf, J., 2020. The future of computing beyond Moore's law. Phil. Trans. R. Soc. A 378 (2166), 61–76.

Son, H., Kim, C., Kim, H., Han, S.H., Kim, M.K., 2010. Trend analysis of research and development on automation and robotics technology in the construction industry. KSCE J. Civ. Eng. 14 (2), 131–139. http://dx.doi.org/10.1007/s12205-010-0131-7.

Song, K., Kim, K., Lee, S., 2017. Discovering new technology opportunities based on patents: Text-mining and F-term analysis. Technovation 60, 1–14.

Steinmetz, N., 2011. Rational Iteration: Complex Analytic Dynamical Systems. Walter de Gruyter.

Sutton, C., Gong, L., 2017. Popularity of arxiv. org within computer science. ArXiv ArXiv:1710.05225.

Tenney, I., Das, D., Pavlick, E., 2019. BERT rediscovers the classical NLP pipeline. ArXiv ArXiv:1905.05950.

Tsvetanov, T., Slaria, S., 2021. The effect of the colonial pipeline shutdown on gasoline prices. Econom. Lett. 209, 110122. http://dx.doi.org/10.1016/j.econlet.2021.110122, URL https://www.sciencedirect.com/science/article/pii/S0165176521003992.

Verhulst, P.-F., 1838. Notice sur la loi que la population suit dans son accroissement. Corresp. Math. Phys. 10, 113–126.

Xi, X., Wei, J., Guo, Y., Duan, W., 2022. Academic collaborations: A recommender framework spanning research interests and network topology. Scientometrics 1–22.

Yang, Y., Liu, Y., Li, H., Yu, B., 2015. Understanding perceived risks in mobile payment acceptance. Ind. Manag. Data Syst. 115 (2), 253–269. http://dx.doi.org/10.1108/IMDS-08-2014-0243.

Yang, J., Sarathy, R., Lee, J., 2016. The effect of product review balance and volume on online shoppers' risk perception and purchase intention. Decis. Support Syst. 89, 66–76. http://dx.doi.org/10.1016/j.dss.2016.06.009.

Yüzügüllü, E., Deason, J.P., 2007. Structuring objectives to facilitate convergence of divergent opinion in hydrogen production decisions. Energy Policy 35 (1), 452–460. http://dx.doi.org/10.1016/j.enpol.2005.12.001.

Zhang, C., Mayr, P., Lu, W., Zhang, Y., 2020a. Extraction and evaluation of knowledge entities from scientific documents: EEKE2020. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. pp. 573–574.

Zhang, Y., Porter, A.L., Cunningham, S., Chiavetta, D., Newman, N., 2020b. Parallel or intersecting lines? Intelligent bibliometrics for investigating the involvement of data science in policy analysis. IEEE Trans. Eng. Manage. 68 (5), 1259–1271.

Zhang, Y., Zhao, R., Wang, Y., Chen, H., Mahmood, A., Zaib, M., Zhang, W.E., Sheng, Q.Z., 2022. Towards employing native information in citation function classification. Scientometrics 1–21.

Zharov, V.S., Kozlov, A.V., 2018. Management of technological development of enterprises on the basis of a life cycle model. In: Quality Management, Transport and Information Security, Information Technologies. IT&QM&IS, IEEE, pp. 181–184.

Dr. Dimitri Percia David is a Professor of Data Science & Econometrics at the University of Applied Sciences of Western Switzerland (HES-SO Valais-Wallis). Prior to this position, he was a postdoctoral researcher at the Information Science Institute of the University of Geneva. He was also the 1st recipient of the *Distinguished CYD Postdoctoral Fellowship*. In his research, Dimitri applies data science and machine learning to the field of technology mining. In 2021, he was the Technical Program Committee chair of *The 16th International Conference on Critical Information Infrastructures Security* (CRITIS 2021) hosted at EPFL. In 2020, he earned his Ph.D. in Information Systems from HEC Lausanne. Prior to that, he worked 8+ years in the commodities-trading industry and as scientific collaborator at ETH Zurich.

Dr. Loïc Maréchal holds a Ph.D. in finance from the University of Neuchâtel and works at HEC Lausanne for the Cyber-Defence Campus, armasuisse, Science and Technology. In this position, he applies private equity models to cybersecurity. He has over ten years of commodity markets experience, which includes working on trading desks and academic research, as well as seven years of lecturing experience. He has a strong interest in machine learning and natural science methods applications for finance.

William Lacube is a cybersecurity researcher at the Cyber-Defence Campus of armasuisse Science and Technology. He is also a cybersecurity researcher at the NATO Cooperative Cyber Defence Centre of Excellence. His main research interests are in the area of telecommunication, cellular networks and Internet of Things.

Sébastien Gillard is a Ph.D. candidate in Information Science at the University of Geneva. He holds a M.Sc. in Physics. He is also a scientific collaborator at ETH Zurich. Sébastien's domains of expertise are Computational Physics and Statistics. As a physicist, he offers a new point of view on socio-technical concepts. His research field is cyber-threat intelligence for cyber-security.

Michael Tsesmelis is a Technology Monitoring researcher at the Cyber-Defence Campus of armasuisse Science and Technology. After graduating from the University of St. Gallen with a B.Sc. in Business Administration, Economics and Data Science, he joined the Swiss Special Forces Command and later the Cyber Battalion of the Swiss Armed Forces. His research focuses on economic data science, computational science and cybersecurity.

Dr. Thomas Maillart is a senior lecturer at the Information Science Institute, University of Geneva, and aims to investigate, model and enhance human collective intelligence through better understanding of incentives, structures and dynamics of social interactions online and in the physical world. In particular, Thomas is interested in the danger and opportunities arising from the fast expanding cyberspace. Thomas holds a Master from EPFL (2005) and a Ph.D. from ETH Zurich (2011). Before joining the University of Geneva, he was a postdoctoral researcher at UC Berkeley until 2016.

Dr. Alain Mermoud is the Head of Technology Monitoring and Forecasting at the armasuisse Science and Technology Cyber-Defence Campus based in the EPFL Innovation Park. His main research interests are emerging technologies, disruptive innovations, (cyber) threat intelligence, and the economics of (cyber) security. In 2021, he was the General chair of the 16th International Conference on Critical Information Infrastructures Security (CRITIS 2021) hosted at EPFL. In 2019, he earned his Ph.D. in Information Systems from HEC Lausanne. Before that, he worked for 5+ years in the banking industry and as a lecturer at ETH Zurich.