Author Accepted Manuscript (DOI: 10.1016/j.techfore.2021.120852), September 2021 – Volume 170 – *Technological Forecasting and Social Change*

Understanding the Long-Term Emergence of Autonomous Vehicles Technologies

Authors:

Seokkyun Woo, School of Public Policy, Georgia Institute of Technology, 685 Cherry St NW, Atlanta, GA 30332 USA. wsk618@gatech.edu, (telephone) +1 404-385-8577 (corresponding author)

Jan Youtie, Enterprise Innovation Institute, Georgia Institute of Technology, 75 Fifth Street, NW Suite 3000, Atlanta, Georgia 30308 USA. jan.youtie@innovate.gatech.edu, (telephone) +1 404-894-6111

Ingrid Ott, Chair in Economic Policy, Karlsruhe Institute of Technology, KIT Campus Süd, Geb. 01.87 - 5. OG, Rüppurrer Str. 1a, Haus B, 76137 Karlsruhe, Germany. ingrid.ott@kit.edu, (telephone) +49 721-608-45257

Fenja Scheu, Chair in Economic Policy, Karlsruhe Institute of Technology, KIT Campus Süd, Geb. 01.87 - 5. OG, Rüppurrer Str. 1a, Haus B, 76137 Karlsruhe, Germany. fenja.scheu@kit.edu, (telephone) +49 721-608-43073

Abstract

Identifying emerging technologies has been of long-standing interest to many scholars and practitioners. Previous studies have introduced methods to capture the concept of emergence from bibliographic records, including the recently proposed Technology Emergence Indicator (Carley et al. 2018). This indicator method has shown to be applicable to various technological fields. However, the indicator uses a limited time window, which can overlook the potential long-term evolution of emerging technologies. Moreover, the existing method suffers from interpretability, because it can be difficult to understand the context in which identified emerging terms are used. In this paper, we propose an improved version of the Technology Emergence Indicator that addresses these issues. In doing so, we examine emerging topics within the field of autonomous vehicles technologies during the period of 1991-2018, guided by a proposition about the long-term diffusion of an emerging technology topic. The results show that different autonomous vehicle technology topics emerge during each of the three 10-year periods under analysis, including an initial period of understanding the surrounding environment and path planning, a second period marked by DARPA Grand Challenge motivated factors associated with the urban environment and communication technologies, and a third period relating to machine learning and object detection. This association with certain emerging technology topics in each decade is also characterized by different trajectories of continued or cyclical carryover across the decades. The results suggest a methodology that practitioners can use in examining research areas to understand which topics are likely to persist into the future.

Keywords: autonomous vehicles, emerging technologies, technology emergence indicator, word-embedding

1. Introduction

Accounts of the history of autonomous vehicle technologies appear to move consistently toward increasing emergence in scientific and engineering communities. Wetmore (2003) discusses how autonomous cars were initially based around intelligence embedded in the road system and with cars equipped with receivers, for example, the General Motors Firebird II. Meanwhile, certain function-specific systems became more automated and were introduced in cruise control and anti-locking brakes. Pilots of various vehicles to test greater autonomy continued. Government funding of academic research into more fully autonomous vehicles took place in the 1970s, 1980s, and 1990s. The DARPA Grand Challenges in 2004, 2005, and especially 2007 demonstrated the use of technologies that can navigate in traffic. Academic research has continued alongside corporate pilot autonomous vehicles and prize competitions that have increasingly used game-based technologies and deep machine learning (Stilgoe 2018).

Unmanned underwater vehicles have a similar technology trajectory, with early pilots focused on data collection advancing to more recent efforts involving government funding, research, and corporate activity around the synthesis of sensors, communication, navigation, artificial intelligence, and energy management (Blidberg 2001, Wynn et al. 2014). Although torpedoes can be considered one of the earliest examples of unmanned underwater vehicles, applications for deep-sea mapping have expanded into deepwater offshore oil and gas production, mineral exploration, and maritime security, among others, as artificial intelligence and energy management improve data gathering, transmission exchange, and wayfinding. Unmanned aerial vehicles also have airborne torpedoes and missiles in their origins, and, although they must fly, they use radio, sensors, computing, software, and mapping to address needs for automatic stabilization, remote control, and navigation.

Some of these autonomous vehicle technologies seemingly follow a progressive technology R&D path in the technology diffusion literature, while others fall by the wayside, such as Nikola Tesla's dirigible-based design for unmanned aerial vehicles (Newcome 2004). Macro- and micro-level factors are involved in the uptake or drop-off of technologies not only in moving to market but also within the R&D domain itself. Observers have commented on the influence of large scale macrosystems on the trajectory of emerging technologies such as autonomous vehicles. Geels and Kemp (2007) highlight the multi-level influences of transitions in socio-technical systems that can facilitate or impede the adoption of technologies. Yun et al. (2016) similarly indicate that causal loops between technology and market result in inconsistencies in the development of autonomous car technologies as represented by patents. Consideration of these systematic, including regulatory, factors varies depending on the level of driver control versus automation, acknowledging that autonomy is subject to gradations rather than being a dichotomous concept (Skeete 2018). At the micro-level, there are concerns about safety, disclosure, and privacy in artificial intelligence databases and psychological needs for confidence and trust (Hengstler, Enkel and Duelli 2016, Hohenberger, Spörrle and Welpe 2017, Penmetsa et al. 2019, Stilgoe 2018). Although some of these factors are more market than R&D oriented, they may well have a role in affecting scientific as well as market emergence of autonomous technologies. For example, scientists are members of the public and can take their public values into the research setting (Scheufele et al. 2007, Youtie et al. 2011).

This paper aims to identify patterns in which technologies used in R&D on autonomous vehicles fall by the wayside or continue to move forward. To this end, we seek to provide information that can help distinguish among models in the literature on the diffusion of innovation and technological emergence. Our presumption is that terms that persist follow one of the models of technology emergence: the standard diffusion of innovation model of continued growth, or the cyclical model of growth, decline, and possible re-emergence. To identify and evaluate such a variety of growth patterns, we propose to apply the Technology Emergence Indicator to a longer time window than the single 10-year period it was originally developed for, that is, to three decades in autonomous vehicle development. We incorporate these multiple observation windows to fully address the evolution of topics. Moreover, we address the difficulties in interpreting isolated emerging terms by comparing groups of closely related terms, which are clustered based on their word-embedded vectors. We show that this greatly improves the interpretability of emerging terms. We focus on autonomous vehicles because of their long history and foundation in and involvement with multiple technologies. Autonomous vehicles are important economically, in addition to from an R&D standpoint.

The analysis based on our improved version of the Technology Emergence Indicator shows that distinctive technologies emerge during each of the three 10-year periods under analysis: (1) an initial period of understanding the surrounding environment and path planning, (2) a second period marked by DARPA Grand Challenge motivated factors associated with the urban environment and communication technologies, and (3) a third period relating to machine learning and object detection. This association with certain emerging technologies in each decade is also characterized by different trajectories of continued or cyclical carryover across the decades.

The paper is organized as follows. In section 2, we discuss previous literature on technological emergence. In section 3, we describe our method for measuring emerging topics and identifying and extracting publication metadata for autonomous vehicle research fields. Section 4 provides our results, including the list of prominent emerging topics that vary across the three periods. Section 5 concludes with discussions and implications for policy and future research.

2. Literature on Technological Emergence

There is a long history of studies about emerging technologies. Much of this literature focuses on defining just what is an emerging technology, particularly the characteristics that make it both emerging and technology. Some of the seminal works are Thomas Kuhn's book about the rise and adoption of paradigms in scientific research (Kuhn 1962), Dosi's technological paradigm, with which he describes discontinuities in technological trajectory as the emergence of new technological paradigm (Dosi 1982), Utterback and Abernathy's introduction of the concept of dominant design (Utterback and Abernathy 1975), and Corning's work based in complexity theory on the self-organization of emerging technologies (Corning 2002). These works are important in that they broadly highlight how a field or technology appears out of a burgeoning research area, emphasizing the dynamic elements of scientific fields.

Recent studies of Rotolo and colleagues and Villaseñor Terán (2017) provide excellent reviews of more contemporary works on emerging technologies. In the case of Rotolo, Hicks and Martin (2015), the literature review extracts from and analyzes the characteristics of 12 studies and isolates a resulting set of five characteristics of emerging technologies: novelty, growth, coherence, impact, and uncertainty. Teran's philosophical review focuses more on the connotations of emerging technologies as a dynamic label for other terms of art: revolutionary technologies, converging technologies, technologies with social impact and/or media attention, and technologies in an early stage of development. Teran discusses the emphasis of works on the "emergent" word in the "emerging technology" phrase and proposes that greater attention also be given to the "technology" part of the phrase, specifically the lack of precision as to whether the technology actually refers to one or more fields or paradigms and whether emerging technologies constitute a single technology or multiple technologies. This proposition harkens to Arthur (2009)'s work about how many new technologies are actually embedded in a system of new and mature technologies.

Both reviews suggest that there is another dimension besides novelty and growth to emerging technologies. Novelty is interesting in itself because of its dynamic nature over time. There is a 50-year history of the trajectory of the study of new technologies, which is based on the work of Everett Rogers's Diffusion of Innovation published in 1962 (Rogers 1962). His work, simply put, proposed a single diffusion curve based on five adopter categories of innovation in the target population.

However, more recent conceptualizations about innovation do not follow a single curve. Other ideas about systems conversation do not follow one curve, in part, because innovations relate to broader research, market, infrastructure, and institutional systems, among others. The Gartner Hype Cycle (Linden and Fenn 2003) and Schmoch (2007)'s double boom concept suggest that technologies are perceived as novel and grow, but then decline, and some re-emerge in a second wave based on a different set of factors including market acceptance (Raffaelli 2019) or technology breakthroughs (Schmoch 2007). Porter and Carley likewise found that emergent technologies do not follow a consistently upward path but have multiple cycles of decline and re-emergence over a 10-year period (Carley et al. (2017)). Novelty is unsatisfactory by itself because coming up with something new per se is necessary but not sufficient for ongoing impact. For example, the idea of the self-driving car is not new. General Motors has been working on it since the 1950s, but early self-driving cars did not work well with the road system (Bimbraw 2015). The notion of emerging technology is that there has to be something else to make the technology endure in addition to novelty.

What is the other axis? Rotolo, Hicks and Martin (2015) suggest candidates such as coherence, impact, and uncertainty. Suggested in the work of Uzzi et al. (2013) is the importance of combining novelty and conventionality (Uzzi et al. 2013). Another set of candidates is proposed by the FUSE project, which, along with novelty and growth, recommends community and persistence (Murdick 2012). We focus on this concept of persistence as an overarching descriptor for technologies that not only have novelty and growth but also endure over a sufficient length of time to be recognized as viable. The general purpose technology literature likewise contends that in addition to breakthroughs, technology must be able to weather the lag time for diffusion to enable it to exhibit the characteristics of a general purpose technology: pervasiveness, innovation spawning, and scope for improvement (Bresnahan and Trajtenberg 1995).

Persistence is envisioned both as a measure of time but also incorporates a sense of dynamism as different factors hold sway. One example of this dynamism is the technology life cycle theory, which argues that in the early stage of a technology (emerging period), there is considerable experimentation and uncertainty. This emerging period is followed by a period of growth where the beginnings of the dominant design appear with increasing standardization, which eventually coalesces and matures. Diversity dominates early growth stages, but standardization becomes more prominent as the technology interacts with external factors, including markets (Dosi 1982, Utterback and Abernathy 1975). Rotolo and colleagues use nanotechnology to exemplify this pattern in which Roco (2004) depicts several generations of nanotechnology, from active nanotechnology to passive nanotechnology, to integrated systems that include nanotechnology.

This section has shown that there is a conceptual discussion of what technology emergence means. There have been measures put forth to identify emerging technologies, including the Technology Emergence Indicator (Carley et al. 2017, Carley et al. 2018, Kwon et al. 2019, Liu and Porter 2020, Ranaei et al. 2020, Wang et al. 2019). This indicator method incorporates the characteristics of emerging technologies emphasized in these papers, such as novelty, growth, and persistence. In the next section, we provide an improved methodology and introduce its application in the field of autonomous vehicle technology.

3. Methods

In this section, we first discuss the literature on measuring technology emergence and introduce the Technology Emergence Indicator (TEI). We then discuss the construction of data and the implementation of the method in detail. Figure 1 presents an overview of the methods we used by illustrating a series of steps by which emerging terms are cleaned, identified, consolidated, and evaluated. We first delineate the field boundary of autonomous vehicle technology with our carefully constructed search terms and obtain bibliographic records related to autonomous vehicle technology. We then parse noun-phrases from the title and abstract fields, which we refer to as “terms” that represent topics within autonomous vehicle technologies. We apply two separate algorithms to these terms to obtain their degree of emergence and their vector representations. The degree of emergence is calculated using TEI (Carley et al. 2018), while each term’s vector space is generated by a word-embedding method from using the Word2Vec algorithm (Mikolov et al. 2013). The corresponding vector space for emerging terms is then used for identifying clusters of similarly located terms. By looking at these clustered terms, we can better understand and evaluate the contexts in which these terms are used in autonomous vehicle technology.

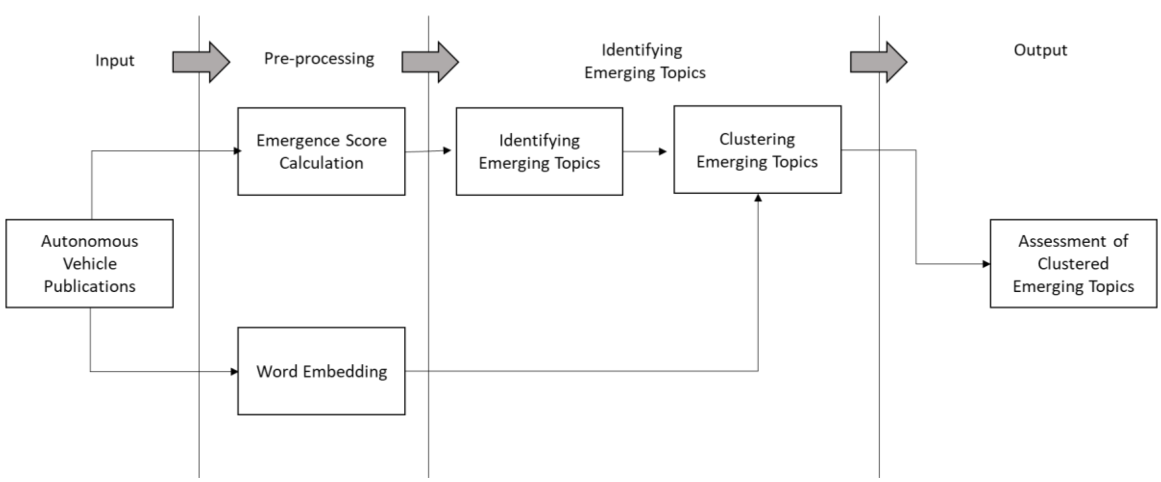

Figure 1. Generating Groups of Emerging Topics.

3.1 Measuring emergence

The lack of a unifying definition of emerging technologies has resulted in previous studies using varying methods to identify technologically emerging topics. For example, Porter et al. (2002) adopted a broad definition of emergence, “technologies that could exert much enhanced economic influence in the coming (roughly) 15-year horizon” (p. 189), and thereby investigated differential national R&D capacity on externally selected emerging topics. Lacasa, Grupp and Schmoch (2003) considered historical development in understanding technological change and innovation in the chemical industry. More recent attempts exploit the text mining approach to identify emerging topics. For example, Glänzel and Thijs (2012) used core documents, a set of documents strongly linked to a given number of documents via co-citation, to identify new emerging topics. Small, Boyack and Klavans (2014) identified clusters of documents that are new and growing rapidly as emerging topics. However, document-level methods can require extensive manual review and interpretation, which is difficult for large-scale datasets. Recently, Carley et al. (2018) came up with a novel text mining method, the TEI algorithm, to identify emerging topics without directly employing clustering methods. This method has several advantages over existing methods. First, it captures emerging topics (noun-phrases) by identifying their novelty, growth, and coherence (a dimension of the community), which comprise the main characteristics of emerging technologies defined by Rotolo, Hicks and Martin (2015). The algorithm also incorporates a persistence element, the importance of which to enduring emerging technologies was raised by Murdick (2012). Therefore, the algorithm is based on a clearly defined definition of emerging technologies. Secondly, the unit of analysis is based on noun-phrases, which can be a good proxy for topics in the context of scientific articles (Milojević 2015). Third, users can select a time frame for analysis. In this aspect, our research draws on Carley et al. (2018)’s TEI algorithm to identify topics within the broad autonomous system technology domain. However, this paper extends the TEI approach by (1) applying the algorithm over multiple 10-year periods, and (2) clustering the emerging terms using the word embedded vectors (Mikolov et al. 2013, Zhang et al. 2018).

3.2 Data

Crafting a viable search strategy for autonomous vehicle technologies can be challenging due to the system-based nature of the technologies. For example, related terms like “smart” and “intelligent” have extensive usage outside of the autonomous vehicle domain, increasing the risk of capturing research that is irrelevant to our interest. Our search strategy is based on the work of Youtie et al. (2017). We began by focusing on system-level technology, not limiting our search to specific, cutting edge technological components. While automation is a frequent keyword and concept that appears in autonomous vehicle publications, we focused on “autonomy” as opposed to terms like “automatic” or “unmanned” that can also refer to remote human operator-controlled systems. We also included terms such as “smart” or “intelligence,” which are indicators of autonomy. We empirically tested these outputs and decided not to include the terms “automated” because that term produces documents related to remote human operator-controlled systems. We prioritized precision over recall, meaning that we sought well-targeted, representative “autonomous systems” datasets to draw accurate conclusions, and sacrificed some portion of relevant data rather than trying to process noisier data.

We constructed sets of keyword-based search queries, which are shown in Table 1. These search terms were first identified by identifying the most common keywords from the initial search, which only include a few terms related to autonomous vehicle technology. We then iteratively refined our search terms. See Youtie et al. (2017) for detail.

Table 1. Search Queries for Autonomous System Technologies

Query 1:	TS = ((Self-driving or autonomous or driverless) near/4 (marine or underwater or transport* or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane))
Query 2:	TS = ((drone near/2 autonomous) or (uav near/4 autonomous) or (auv near/4 autonomous))
Query 3:	TS = ((robot* near/1 (underwater or marine or transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) AND (autonomous or self-driving or driverless))
Query 4:	TS = (“autonomous driv*”)
Query 5:	TS = (((robot* near/1 (underwater or marine or transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) OR (drone or uav or auv)) AND (path or planning or planner or plan))
Query 6:	TS = (((robot* near/1 (underwater or marine or transport* or mobile or car or motorcar or vehicle or automobile or aircraft or airplane or aeroplane)) OR (drone or uav or auv)) AND (2D or 2-D or 3D or 3-D or map or localization or tracking or navigat* or obstacle or avoid*)))

Note: A field “TS” includes title, abstracts and keywords from records.

Our data comes from the Web of Science Core Collection (WOS), which provides detailed metadata information for more than 73 million research articles published in over 20,000 journals. The metadata includes detailed information about titles and abstracts as well as other fields. Our search (Table 2) yielded (as of December 2, 2018) 53,678 publications. Since our objective is to identify emerging topics, we removed publications that were not original articles or proceedings, such as reviews, editorials, and corrections. This left us with 53,005 publications from the years 1956 to 2018.

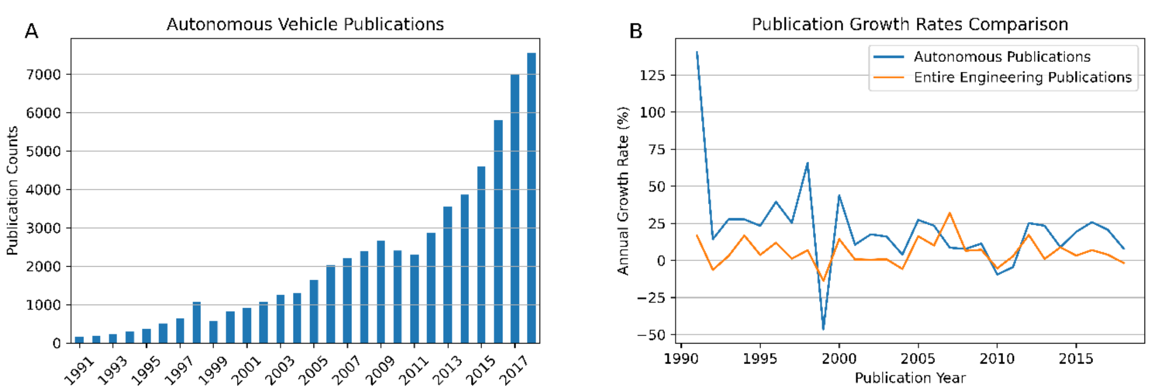

Figure 2. Growth of AVT Publications (A) and its Growth Rates Comparison with Engineering Publications (B)

Figure 2 shows annual publications related to autonomous vehicle technology. Overall the annual publication counts exhibit an increasing upward trend as expected from the general perception of this technology. For example, the number of peer-reviewed publications from 2000-2009 is nearly three times the number of articles from 1991-2000. The publications from the recent 10-year period (2008-2017) are more than twice the number of publications from 2000-2009.

Table 2. Comparison of the Three Periods

	Period 1	Period 2	Period 3
Period	1991-2000	2000-2009	2009-2018
Average yearly growth rate (%)	25.00	14.11	14.36
Total Records	4,760	16,273	42,432

At the same time, there seem to be some fits and starts in the annual counts of publications against this overall growth trend. The first divergence from consistent year-over-year growth is in the late 90s dot-com bubble, after which time a decline in publications is observed. We then see a second wave of growth until the Great Recession, after which another decline in annual publication counts appears. More recently, the field has been experiencing rapid growth from the early 2010s, which may be led by recent expansion in the artificial intelligence research frontier. We take this occurring pattern as a reason to identify emerging topics for three distinct periods of time. We define the first period from 1991 to 2000, the second period from 2000 to 2009, and the third period from 2008 to 2017. Notice that we allow some overlaps between the periods to account for the continuous nature of this technology's development. Table 3 shows select differences between the periods. In terms of the growth rate, the first period saw the most rapid growth rate, followed by the second and the recent periods. However, the most recent period includes years just before the Great Recession, which may have the effect of attenuating the field's growth rate. When we consider records from 2010, the last period's average growth rate is 16%, which is higher than the average growth rate from the second period.

3.3 Identifying emerging terms

We describe the TEI algorithm step by step, which has been applied and tested against numerous scientific fields (Carley et al. 2018, Huang et al. 2018, Kwon et al. 2019, Ranaei et al. 2020, Wang et al. 2019). The algorithm's objective is to identify a set of terms (noun-phrases) from the records that satisfy the conditions put forth by the definition of emerging technology and calculate each term's respective emerging scores. We first extracted the terms - which are our potential emerging topics - from the titles and abstracts of the identified publications. We began this extraction by merging the title and abstract fields for each publication. We then extracted every noun-phrase from these fields[1]. We removed stop words and common academic terms in the process. Then we applied five standard thesauri in "Refine NLP" algorithm[2] from VantagePoint software (thevantagepoint.com) to further consolidate the noun phrase terms and remove any stop words or multi-word phrases. This process resulted in 98,410 unique terms over the complete time frame (1991-2018).

3.3.1 Identifying candidate terms

We now describe the TEI algorithm step by step, which has been applied and tested against numerous scientific fields (Carley et al. 2018, Huang et al. 2018, Kwon et al. 2019, Wang et al. 2019). The TEI algorithm requires terms from publications over a ten-year period as input.[3] For each of the three periods, we separate the publications into a base period (first three years) and an active period (later seven years). The reason for this separation is that the TEI algorithm makes a direct comparison of the relative term frequencies between the base and active periods to calculate novelty and growth while ensuring that a term has some degree of persistence over the period. Table A.1 in Appendix A illustrates every step for the calculation and the corresponding mathematical equations.

Steps I to IV are applied to identify terms that satisfy (I) persistence, (II) novelty and growth, (III) community, and (IV) scope criteria. Term persistence ensures that a term appears in at least three years over the ten-year period and appears in at least seven publications in the active period. The novelty criterion ensures that a term cannot appear in as many as 15% of the base period publications so that an emerging term is relatively rare in the earlier period. The growth criterion ensures that the growth of the publications using emerging terms should be at least 1.5 times greater than the general growth of the field. The c*ommunity* criterion in Step III ensures that emerging terms are used by more than a single organization to avoid the possibility that the usage of a term is driven by a highly productive but isolated research group. *Scope* in Step IV is used to filter out extremely common terms in the corpus. Candidate emerging terms are those that satisfy the filtering criteria laid out in Step I through Step IV, with additional common terms such as from the search query having been manually removed. The sensitivity of the parameters used in the e-score has been tested by Liu and Porter (2020); the tests found them to be relatively stable.

3.3.2 E-score calculation

For every candidate term, we calculate their corresponding e-score using the TEI algorithm. Step V in Table A.1 in Appendix A shows the formula of the calculation. The e-score is a weighted sum of three normalized differences over a 10-period timeframe: (1) twice the difference between the 8th through the 10th

---

[1] There are few open-source NLP packages (Spacy or NLTK) that can readily be used for extracting noun-phrases from sentences. However, we use VantagePoint software as they provide a finely tuned algorithm, particularly suitable for extracting non-perfunctory noun-phrases from academic documents.

[2] ClusterSuite consolidates VantagePoint's five standard thesauri along with various fuzzy matching and other cleaning and text consolidation routines. ClusterSuite is developed by J.J. O'Brien (O'Brien et al., n.d.) and available at www.VPInstitute.org.

[3] In fact, the time window does not have to be limited to 10 years. We can choose any time window, but we believe that 10 years is reasonable window to observe evolutions of topics.

periods and the 4th through the 6th periods; (2) ten times the difference between the 9th and 10th period, and the 7th and 8th period; and (3) ten times the average annual growth rate using the difference between the 10th period and the 7th period divided by three. Once each candidate term's e-score is calculated, we regard terms with a non-negative score above a certain threshold as emerging topics.

3.4 Clustering individual emerging concepts

Although the current method allows for identifying emerging terms, it can be difficult to make sense of them because we do not know the contexts in which they are used. One way to address this issue is to cluster similarly used terms together to have a better sense of how these terms are used. Clustering terms are based on their vector representation. A conventional method is to construct a term vector from its bag-of-word representation, but this method neglects rich information embedded at a sentence level. Hence, we apply word-embedding (Mikolov et al. 2013) to construct vector representations of words, which are subsequently used for the clustering, which will be discussed in detail in the next section.

4. Results

4.1 Top emerging terms from autonomous vehicle technology

Table 3 reports the lists of top emerging terms separately calculated from the 1991-2000, 2000-2009, and 2009-2018 periods. The top emerging terms from Period 1 are dominated by terms related to research programs on understanding surrounding environments. These include “unknown environment,” “indoor environment,” “GPS,” and “recognition”. We also observe terms related to path planning algorithm research: “localization,” “map building” and “dynamic model.” The top emerging terms from Period 2 includes some terms from the first period related to the path planning algorithm, for example, “grid map” and “mobile robot path planning.” This period also witnessed an essential milestone in the advancement of autonomous vehicle research (Beiker 2014). From 2004, DARPA hosted the three “Grand Challenge” competitions on autonomous vehicles in 2004, 2005, and 2007. While the competitions took place in open deserts in 2004 and 2005, the 2007 Grand Challenge took place in mock urban environments. The top emerging terms from this period (2000-2009) reflect the significance of the research developed as part of the Grand Challenge competitions. The top terms from this period include “urban environments” and “urban challenge”. This period also includes new streams of research programs from communication technologies (“wireless sensor networks”) and signal processing algorithms (“particle filter”).

Table 3. Top 10 Emerging Terms from the Three Periods

Period 1 (1991-2000)		Period 2 (2000-2009)		Period 3 (2009-2018)	
Term	e-score	Term	e-score	Term	e-score
unknown environment	11.99	urban environments	9.53	deep learning	23.54
localization	6.76	dynamic model	6.66	LIDAR	20.98
indoor environment	6.42	swarm robotics	5.44	convolutional neural network	15.22
map building	6.09	urban challenge	4.94	object detection	15.18
surveillance	5.49	wireless sensor networks	4.69	CNN*	13.86
GPS*	5.09	simulation environment	4.15	point clouds	13.35
multiple robots	4.88	particle filter	3.49	smart cities	12.96
tracking problem	4.56	WSN*	3.47	artificial intelligence	12.64
dynamic model	4.47	grid map	3.44	deep neural networks	11.87
recognition	4.18	mobile robot path planning	3.40	external disturbances	11.76

Note: GPS*=global positioning system, WSN* = wireless sensor network, CNN* = convolutional neural network

The top emerging terms from Period 3 contrasts sharply with the emerging terms of the preceding periods. Unlike the previous periods, terms related to path planning algorithms are not listed among the most common terms in Period 3. Instead, terms related to machine learning dominate the list. These are terms include “deep learning,” “convolutional neural network,” “artificial intelligence,” and “deep neural networks.” We also see terms related to object detection research, including “LIDAR,” “point clouds,” and “object detection.” Policy-driven research such as “smart cities” was also in the top list.

4.2 Patterns of emerging terms

The list of emerging terms in Table 3 allows us to understand what technology is emerging at a given period, but it does not provide the specific context in which emerging terms are used. The ability of the TEI algorithm to identify hundreds of emerging terms makes it difficult to explore how these emerging terms are related to each other. It is thus useful to identify groups of emerging terms that are similar in concept. There are several ways of clustering emerging terms. The traditional approach is to calculate a bag-of-words vector for each record from the raw text in the title and abstract fields. Then, various clustering techniques are applied to identify topics that are more likely to co-appear in the record. The bag-of-words method can be unwieldy because of its large feature dimensions and sparse representation.

4.2.1 Vector representation of emerging terms

Recent advances in natural language processing allow us to represent the relation of words not based on the sparse frequency networks but as vectors in dense high dimensional space (Mikolov et al. 2013). Collectively known as word embedding, this new vector space model has attracted widespread attention from not only computer scientists but increasingly from social scientists (Gentzkow, Kelly and Taddy 2019, Kozlowski, Taddy and Evans 2019) as well. For our research, there are two ways to construct a vector space model using word embedding methods. We could rely on a pre-trained word embedding model. However, we believe that a pre-trained model would provide an accurate representation of words only in a general context. Because terms tend to be field-specific in academia, we contend that a training model using our dataset would provide a more accurate representation of terms in the context of autonomous vehicle technology (this approach was also used by Zhang et al. (2018)).

We trained our word embedding model on our autonomous dataset with the Word2Vec algorithm[4]. The trained word embedding model assigns a dense vector for each word. Some of our emerging terms include more than one word (i.e., "machine learning"), making the assignment of word embedding vectors to each word problematic. We assigned the term frequency-inverse document frequency (Tf-IDF) weighted sum of the word vectors to these terms to address this problem, thereby designating greater weights to less frequent term vectors. We also used relative frequency weight, in which each term would be assigned with weight associated with how common a word is in the corpus relative to its neighboring word in a phrase, as part of the validation effort. Clustering, based on this method, led to consistent results with insubstantial variations (see Appendix C).

4.2.2 Clustering emerging terms

After assigning vectors to each of the emerging terms, we used the k-mean clustering algorithm to assign similarly represented emerging terms into groups of clustered topics. We reported our results using a 30 cluster solution. We did additional analyses using different numbers of clusters (k=40, k=50), which yielded very similar results (see Appendix B for solutions based on alternative numbers of clusters).

Figure 3 illustrates two clustered groups of related emerging terms, plotting the annual changes in the fraction of terms appearing in our dataset from 1991 to 2018. Figure 3(A) shows a group of emerging terms associated with LIDAR (light detection and ranging) technology. The related emerging terms grouped with "LIDAR" include terms directly related to LIDAR (LIDAR data or LIDAR sensor) or to various detection technologies such as "radar sensors" or "light detection." One advantage of using the word-embedding over bag-of-words method is that embedded terms can be clustered together despite having never appeared in the same records. For example, the term "radar sensors" in Figure 3(A) never co-appeared with "LIDAR data". However, these two terms are related in the sense that these terms are often used in similar contexts in autonomous vehicle technology research. Emerging terms in Figure 3(B) are also very much related to machine learning and artificial intelligence research. Appendix B groups topics using k=40 (Figure B.1) and k=50 (Figure B.2) solutions. While Figure 3(B) groups AI and CNN/Deep

[4] We used python Gensim package to implement Word2Vec algorithm. (https://radimrehurek.com/gensim/models/word2vec.html)

learning into the same cluster, when we allow for more topical solutions (i.e., higher values of k), we see that these two groups are into two different clusters.

4.2.3 Identifying varying patterns of emerging terms

Prominent emerging topics can monotonically persist over time. An example of this is shown in Figure 3(A), which represents clustered terms related to LIDAR, which is an essential component of autonomous vehicle technologies. LIDAR illuminates targets with a laser and measures reflected light with sensors, which then can be used to generate 3D environment maps (Schwarz 2010). These LIDAR functions allow autonomous vehicles to detect and avoid obstacles. First used by meteorologists and cartographers, LIDAR became one of the crucial components of autonomous vehicles when high-definition rotating head LIDAR (i.e., 360-degree image information multiple sensors generating millions of point data in second) was introduced during the DARPA Grand Challenges. The rapid growth of the fraction of publications on the LIDAR topic from the late mid-2000s is consistent with the innovation and application of LIDAR technology to autonomous vehicle research during and after the DARPA Grand Challenges.

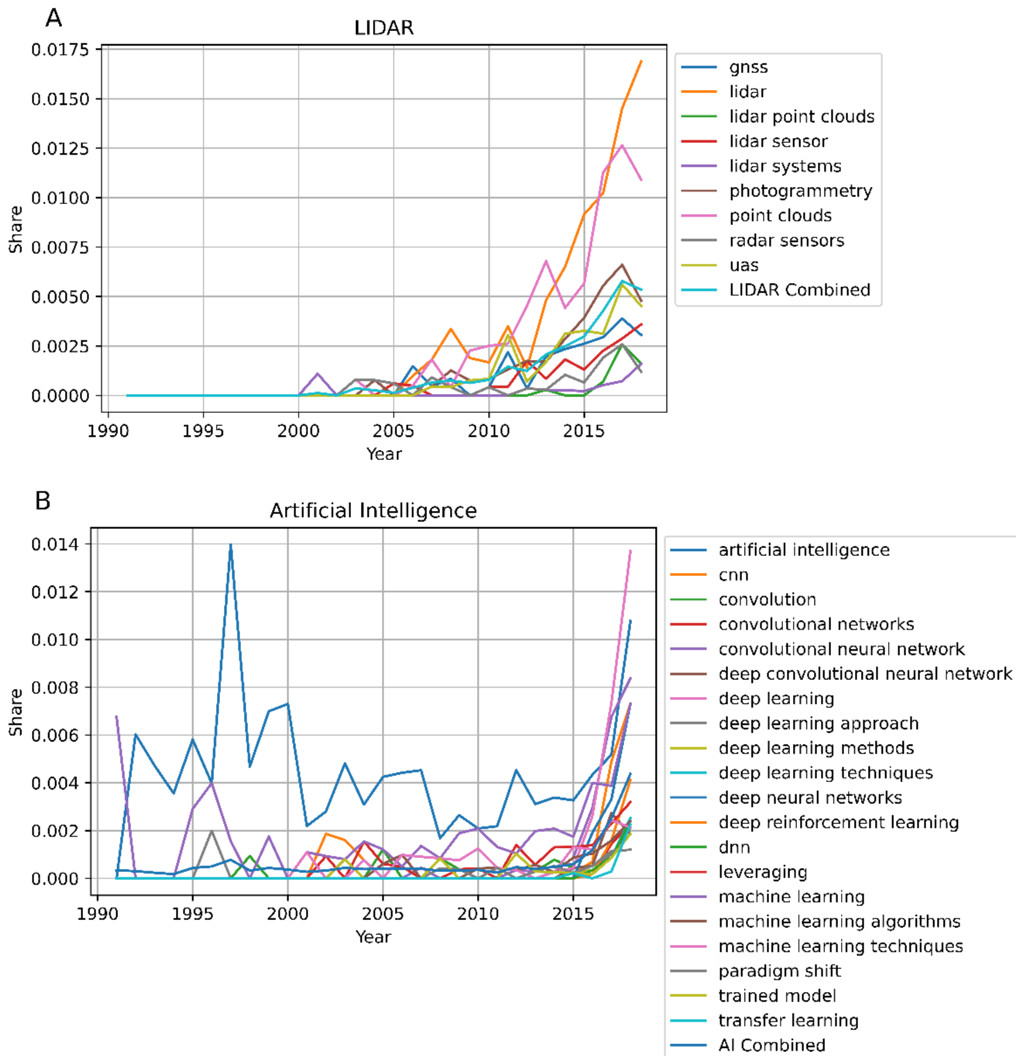

Figure 1. Relative Growth in LIDAR (A) and AI (B) topics. The y-axis is the share of publications in a given year in which an emerging term appears. Emerging terms are clustered around a topic using the k-mean clustering method with k=30.

Not all technologies experience linear monotonic growth. Another pattern that can be observed is the U-pattern, where the technology becomes prominent, declines, then re-emerges. The artificial intelligence topic shown in Figure 3(B) follows this pattern. Figure 3(B) shows that machine learning, artificial intelligence, and computer vision start with a relatively high share in the early 1990s. From the mid to late 1990s, their share drops and remains stagnant until the mid-2010s, when they start to experience new rapid growth. We can also see that there are fluctuations and many peaks, although the overall trend follows a U-shape growth pattern.

This trend is consistent with the so-called "AI Winter" (Grosz 2005, Grudin 2009), a period from the late 1980s to the early 1990s, which witnessed reduced funding and interest in AI research. The re-emergence of AI and machine learning research programs within the autonomous vehicle domain – also known as the “AI Spring” (Cruz and Treisman 2018) – have coincided with increasing market demand for AI research as well as recent advancements in computational power and more efficient algorithms (O'Leary 2013). In this aspect, the growth patterns of machine learning research resemble Schmoch (2007)’s double boom concept, which suggests that technologies are perceived as novel and grow, but then decline, and some re-emerge in a second wave based on a different set of factors, including market acceptance (Schmoch 2007).

5. Discussion and Conclusion

This study used and extended the TEI algorithm (Carley et al. 2017) to understand the emergence of autonomous vehicle technologies over the three decades from 1991 to 2018. Our results showed that the emerging terms in the initial 10-year period included associated technologies designed for understanding the surrounding environment and path planning. Those associated with the second decade were marked by DARPA Grand Challenge related terms about the urban environment and new streams of research programs from communication technologies and signal processing algorithms. Terms associated with technologies in the third decade relate to machine learning and object detection. Although each decade appears to be associated with distinct emerging technologies, we do see different patterns of carryover from decade to decade.

Our research makes both conceptual and methodological contributions to the literature on emerging technologies. From a conceptual standpoint, our study suggests that technology emergence is not only about novelty and growth as studies on publication growth “bursts” underscore (Chen 2006). It is also important to look at technology patterns over time to understand the extent to which terms related to these technologies persist in the research literature. Our study showed that some terms persist while others fall away over time. The literature related to technology emergence distinguishes these two dynamics as the diffusion of innovation camp in which technology emerges and continues to grow in an S-curve pattern, and the Uli Schmoch (and Gartner Hype Cycle) camp in which terms demonstrate initial growth followed by a decline then the second period of escalation, indicative of a U-shaped pattern, as market forces or new technology breakthroughs catch up with an early technological promise. Our study shows evidence of both trends. LIDAR technologies exhibit monotonic growth indicative of an S-curve while AI-related technologies exhibit a U-shaped pattern with growth in the first 10 years, then decline and subsequent reappearance in the third 10 years as a result of the growth of machine learning. Both patterns are observed for different technologies within the autonomous vehicle domain.

From a methodological standpoint, our paper uses a replicable quantitative method, which can complement the use of qualitative case studies or anecdotal evidence (Linden and Fenn 2003, Raffaelli 2019). Our paper demonstrates the application and extension of the TEI algorithm to autonomous vehicles technology abstract publication records to uncover emerging technologies. The existing TEI method uses a 10-year time window to evaluate extracted emerging terms (Carley et al. 2018, Huang et al. 2018, Kwon et al. 2019, Wang et al. 2019). This paper demonstrates the benefits of using multiple timeframes to better understand how the saliency of technologies evolves over time. We also performed a sensitivity test using 5 half-year time slicing and generally saw the results were consistent with the main findings, although missing data in the publication month field greatly reduced our sample. The paper also provides a use case of applying a word-embedding algorithm to extract dense vectors that are then used to cluster emerging terms that appear in similar contexts. This greatly enhances our understanding of emerging technologies. While clustering embedded word vectors have been used on the “front end” to consolidate synonyms into a single word (Zhang et al. 2018), we use this method on the “back end” with a different goal, which is to find words that are used in similar contexts. As seen from our examples, providing groups of related emerging terms greatly enhances the interpretability of emerging technologies.

Our analysis is subject to several limitations that should be noted. The e-score takes into account document information and term level information from title and abstract fields. It is not based on the entire full-text of the document, as was used by Small, Boyack and Klavans (2014), who manually read the full-text as part of their cluster assignment approach. We contend, similar to the argument put forth by Carley et al. (2018) and Milojević (2015), that the use of titles and abstracts is likely to be sufficient for picking up the most important terms and it reduces noise. We also acknowledge that this is a study of a single technology domain—autonomous vehicles. The robustness of our methodology and findings of varying patterns of emergence would be strengthened by future research that applies it to other technology domains as well as to different time periods. This paper does not address any mechanism with which ideas emerge or remerge overtime, however which is great interests to many scholars. There have been increasing attempts to address this question using networks of scientific concepts (Foster, Rzhetsky and Evans 2015, Iacopini, Milojević and Latora 2018, Shi and Evans 2019). Future research on the E-score could not only focus on high scoring terms (high velocity) but also look at the changing technology topology constructed from networks of emerging terms (either using co-occurrence or word embedded vectors).

This study has implications for R&D managers in their efforts to use existing data to quantitatively understand where technology might be going. Existing approaches often require extensive quantitative and programming expertise, which may not be readily accessible to many R&D managers. Some of these tasks can be readily automated, such as the acquisition of abstract records, and parsing, cleaning, and scoring them. Our study made use of VantagePoint for these tasks and supplemented them with quantitative programming associated with the use of word embedding for cluster analysis. Dimensional analysis can be difficult to automate and can require some human intervention for assessing the interpretability of the results. This combination of “push button” (through VantagePoint) processing of front-end term extraction, cleaning, and grouping tasks, and programming methods with human interpretation for back-end dimensional analysis tasks may not be readily usable to some R&D managers but represents a movement away from requiring computing expertise for all tasks.

It will continue to be important for R&D managers to be able to apply systematic methods for looking at where to place future R&D investments. Past experience, contact with experts in the field, simulations, and other foresight methods can be helpful (Popper 2008). But many managers have to deal with bounded rationality problems such as limitations on the knowledge they can process in addition to their capacity to automate data acquisition and analysis tasks (Simon 1947). This analysis suggests a mixed automated-computing approach for evaluating R&D investment options in a broad technology domain that not only identifies growing topics but also focuses on those that are likely to persist into the future.

Acknowledgments

The authors thank David Bälz for his assistance in providing data and helpful interpretations. This research has been conducted with supports from the US National Science Foundation (Award #1759960 – "Indicators of Technological Emergence"), and an International Collaboration Grant from the Karlsruhe Institute of Technology. The findings and observations contained herein are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Karlsruhe Institute of Technology.

References

Arthur, W Brian. 2009. *The Nature of Technology: What It Is and How It Evolves*: Simon and Schuster.

Beiker, Sven. 2014. "History and Status of Automated Driving in the United States." Pp. 61-70 in *Road Vehicle Automation*: Springer.

Bimbraw, Keshav. 2015. "Autonomous Cars: Past, Present and Future a Review of the Developments in the Last Century, the Present Scenario and the Expected Future of Autonomous Vehicle Technology." Pp. 191-98 in *2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, Vol. 1: IEEE.

Blidberg, D Richard. 2001. "The Development of Autonomous Underwater Vehicles (Auv); a Brief Summary." Pp. 1 in *Ieee Icra*, Vol. 4.

Bresnahan, Timothy F and Manuel Trajtenberg. 1995. "General Purpose Technologies 'Engines of Growth'?". *Journal of econometrics* 65(1):83-108. doi: https://doi.org/10.1016/0304-4076(94)01598-T.

Carley, Stephen F, Nils C Newman, Alan L Porter and Jon G Garner. 2017. "A Measure of Staying Power: Is the Persistence of Emergent Concepts More Significantly Influenced by Technical Domain or Scale?". *Scientometrics* 111(3):2077-87. doi: https://doi.org/10.1007/s11192-017-2342-x.

Carley, Stephen F, Nils C Newman, Alan L Porter and Jon G Garner. 2018. "An Indicator of Technical Emergence." *Scientometrics* 115(1):35-49. doi: https://doi.org/10.1007/s11192-018-2654-5.

Chen, Chaomei. 2006. "Citespace Ii: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature." *Journal of the American Society for Information Science and Technology* 57(3):359-77. doi: https://doi.org/10.1002/asi.20317.

Corning, Peter A. 2002. "The Re-Emergence of "Emergence": A Venerable Concept in Search of a Theory." *Complexity* 7(6):18-30. doi: https://doi.org/10.1002/cplx.10043.

Cruz, Lito Perez and David Treisman. 2018. "Making Ai Great Again: Keeping the Ai Spring." *IJCCI*.

Dosi, Giovanni. 1982. "Technological Paradigms and Technological Trajectories." *Research Policy* 2(3):l47-62. doi: https://doi.org/10.1016/0048-7333(82)90016-6.

Foster, Jacob G, Andrey Rzhetsky and James A Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80(5):875-908. doi: https://doi.org/10.1177%2F0003122415601618.

Geels, Frank W and René Kemp. 2007. "Dynamics in Socio-Technical Systems: Typology of Change Processes and Contrasting Case Studies." *Technology in society* 29(4):441-55. doi: https://doi.org/10.1016/j.techsoc.2007.08.009.

Gentzkow, Matthew, Bryan Kelly and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57(3):535-74. doi: https://doi.org/10.1257/jel.20181020.

Glänzel, Wolfgang and Bart Thijs. 2012. "Using 'Core Documents' for Detecting and Labelling New Emerging Topics." *Scientometrics* 91(2):399-416. doi: https://doi.org/10.1007/s11192-011-0591-7.

Grosz, Barbara J. 2005. "Whither Ai: Identity Challenges of 1993-95." *Ai Magazine* 26(4):42-42. doi: https://doi.org/10.1609/aimag.v26i4.1844.

Grudin, Jonathan. 2009. "Ai and Hci: Two Fields Divided by a Common Focus." *Ai Magazine* 30(4):48-48. doi: https://doi.org/10.1609/aimag.v30i4.2271.

Hengstler, Monika, Ellen Enkel and Selina Duelli. 2016. "Applied Artificial Intelligence and Trust—the Case of Autonomous Vehicles and Medical Assistance Devices." *Technological Forecasting and Social Change* 105:105-20. doi: https://doi.org/10.1016/j.techfore.2015.12.014.

Hohenberger, Christoph, Matthias Spörrle and Isabell M Welpe. 2017. "Not Fearless, but Self-Enhanced: The Effects of Anxiety on the Willingness to Use Autonomous Cars Depend on Individual Levels

of Self-Enhancement." *Technological Forecasting and Social Change* 116:40-52. doi: https://doi.org/10.1016/j.techfore.2016.11.011.
Huang, Ying, Alan L Porter, Scott W Cunningham, Douglas KR Robinson, Jianhua Liu and Donghua Zhu. 2018. "A Technology Delivery System for Characterizing the Supply Side of Technology Emergence: Illustrated for Big Data & Analytics." *Technological Forecasting and Social Change* 130:165-76. doi: https://doi.org/10.1016/j.techfore.2017.09.012.
Iacopini, Iacopo, Staša Milojević and Vito Latora. 2018. "Network Dynamics of Innovation Processes." *Physical review letters* 120(4):048301. doi: https://doi.org/10.1103/PhysRevLett.120.048301.
Kozlowski, Austin C, Matt Taddy and James A Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905-49. doi: https://doi.org/10.1177%2F0003122419877135.
Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*: University of Chicago press.
Kwon, Seokbeom, Xiaoyu Liu, Alan L Porter and Jan Youtie. 2019. "Research Addressing Emerging Technological Ideas Has Greater Scientific Impact." *Research Policy* 48(9):103834. doi: https://doi.org/10.1016/j.respol.2019.103834.
Lacasa, Icíar Dominguez, Hariolf Grupp and Ulrich Schmoch. 2003. "Tracing Technological Change over Long Periods in Germany in Chemicals Using Patent Statistics." *Scientometrics* 57(2):175-95. doi: https://doi.org/10.1023/a:1024133517484.
Linden, A and J Fenn. 2003. "Understanding Gartner's Hype Cycles: Strategic Analysis Report No." Vol.: R-20-1971. Stamford, CT: Gartner, Inc.
Liu, Xiaoyu and Alan L Porter. 2020. "A 3-Dimensional Analysis for Evaluating Technology Emergence Indicators." *Scientometrics*:1-29. doi: https://doi.org/10.1007/s11192-020-03432-6.
Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." Pp. 3111-19 in *Advances in neural information processing systems*.
Milojević, Staša. 2015. "Quantifying the Cognitive Extent of Science." *Journal of Informetrics* 9(4):962-73. doi: https://doi.org/10.1016/j.joi.2015.10.005.
Murdick, Dewey. 2012. "Finding Patterns of Emergence in Science and Technology." Vol.: OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE WASHINGTON DC INTELLIGENCE ….
Newcome, Laurence R. 2004. *Unmanned Aviation: A Brief History of Unmanned Aerial Vehicles*: American Institute of Aeronautics and Astronautics.
O'Leary, Daniel E. 2013. "Artificial Intelligence and Big Data." *IEEE intelligent systems* 28(2):96-99. doi: https://doi.org/10.1109/MIS.2013.39.
Penmetsa, Praveena, Emmanuel Kofi Adanu, Dustin Wood, Teng Wang and Steven L Jones. 2019. "Perceptions and Expectations of Autonomous Vehicles–a Snapshot of Vulnerable Road User Opinion." *Technological Forecasting and Social Change* 143:9-13. doi: https://doi.org/10.1016/j.techfore.2019.02.010.
Popper, Rafael. 2008. "How Are Foresight Methods Selected?". *Foresight-The journal of future studies, strategic thinking and policy* 10(6):62-89. doi: https://doi.org/10.1108/14636680810918586.
Porter, Alan L, J David Roessner, Xiao-Yin Jin and Nils C Newman. 2002. "Measuring National 'Emerging Technology'capabilities." *Science and Public Policy* 29(3):189-200. doi: https://doi.org/10.3152/147154302781781001.
Raffaelli, Ryan. 2019. "Technology Re-emergence: Creating New Value for Old Technologies in Swiss Mechanical Watchmaking, 1970–2008." *Administrative science quarterly* 64(3):576-618. doi: https://doi.org/10.1177%2F0001839218778505.
Ranaei, Samira, Arho Suominen, Alan Porter and Stephen Carley. 2020. "Evaluating Technological Emergence Using Text Analytics: Two Case Technologies and Three Approaches." *Scientometrics* 122(1):215-47. doi: https://doi.org/10.1007/s11192-019-03275-w.

Roco, Mihail C. 2004. "The Us National Nanotechnology Initiative after 3 Years (2001–2003)." *Journal of Nanoparticle Research* 6(1):1-10. doi: https://doi.org/10.1023/B:NANO.0000023243.25838.73.
Rogers, Everett M. 1962. *Diffusion of Innovations*: Simon and Schuster.
Rotolo, Daniele, Diana Hicks and Ben R Martin. 2015. "What Is an Emerging Technology?". *Research Policy* 44(10):1827-43. doi: https://doi.org/10.1016/j.respol.2015.06.006.
Scheufele, Dietram A, Elizabeth A Corley, Sharon Dunwoody, Tsung-Jen Shih, Elliott Hillback and David H Guston. 2007. "Scientists Worry About Some Risks More Than the Public." *Nature nanotechnology* 2(12):732-34. doi: https://doi.org/10.1038/nnano.2007.392.
Schmoch, Ulrich. 2007. "Double-Boom Cycles and the Comeback of Science-Push and Market-Pull." *Research Policy* 36(7):1000-15. doi: https://doi.org/10.1016/j.respol.2006.11.008.
Schwarz, Brent. 2010. "Lidar: Mapping the World in 3d." *Nature Photonics* 4(7):429. doi: https://doi.org/10.1038/nphoton.2010.148.
Shi, Feng and James Evans. 2019. "Science and Technology Advance through Surprise." *arXiv preprint arXiv:1910.09370*.
Simon, Herbert A. 1947. *Administrative Behavior*: Simon and Schuster.
Skeete, Jean-Paul. 2018. "Level 5 Autonomy: The New Face of Disruption in Road Transport." *Technological Forecasting and Social Change* 134:22-34. doi: https://doi.org/10.1016/j.techfore.2018.05.003.
Small, Henry, Kevin W Boyack and Richard Klavans. 2014. "Identifying Emerging Topics in Science and Technology." *Research Policy* 43(8):1450-67. doi: https://doi.org/10.1016/j.respol.2014.02.005.
Stilgoe, Jack. 2018. "Machine Learning, Social Learning and the Governance of Self-Driving Cars." *Social Studies of Science* 48(1):25-56. doi: https://doi.org/10.1177%2F0306312717741687.
Utterback, James M and William J Abernathy. 1975. "A Dynamic Model of Process and Product Innovation." *Omega* 3(6):639-56.
Uzzi, Brian, Satyam Mukherjee, Michael Stringer and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342(6157):468-72. doi: https://doi.org/10.1126/science.1240474.
Villaseñor Terán, Mauricio. 2017. "Philosophical Explorations for a Concept of Emerging Technologies." *Techné: Research in Philosophy and Technology*. doi: https://doi.org/10.5840/techne201792170.
Wang, Zhinan, Alan L Porter, Xuefeng Wang and Stephen Carley. 2019. "An Approach to Identify Emergent Topics of Technological Convergence: A Case Study for 3d Printing." *Technological Forecasting and Social Change* 146:723-32. doi: https://doi.org/10.1016/j.techfore.2018.12.015.
Wetmore, J. 2003. "Driving the Dream. The History and Motivations Behind 60 Years of Automated Highway Systems in America." *Automotive History Review* 7:4-19.
Wynn, Russell B, Veerle AI Huvenne, Timothy P Le Bas, Bramley J Murton, Douglas P Connelly, Brian J Bett, Henry A Ruhl, Kirsty J Morris, Jeffrey Peakall and Daniel R Parsons. 2014. "Autonomous Underwater Vehicles (Auvs): Their Past, Present and Future Contributions to the Advancement of Marine Geoscience." *Marine Geology* 352:451-68. doi: https://doi.org/10.1016/j.margeo.2014.03.012.
Youtie, Jan, Stephen Carley, Philip Shapira, Elizabeth A Corley and Dietram A Scheufele. 2011. "Perceptions and Actions: Relationships of Views on Risk with Citation Actions of Nanotechnology Scientists." *Research Evaluation* 20(5):377-88. doi: https://doi.org/10.3152/095820211X13176484436014.
Youtie, Jan, Alan L Porter, Philip Shapira, Seokkyun Woo and Yayun Huang. 2017. "Autonomous Systems: A Bibliometric and Patent Analysis." Vol.: Studien zum deutschen Innovationssystem.
Yun, JinHyo Joseph, DongKyu Won, EuiSeob Jeong, KyungBae Park, JeongHo Yang and JiYoung Park. 2016. "The Relationship between Technology, Business Model, and Market in Autonomous Car and Intelligent Robot Industries." *Technological Forecasting and Social Change* 103:142-55. doi: https://doi.org/10.1016/j.techfore.2015.11.016.

Zhang, Yi, Jie Lu, Feng Liu, Qian Liu, Alan Porter, Hongshu Chen and Guangquan Zhang. 2018. "Does Deep Learning Help Topic Extraction? A Kernel K-Means Clustering Method with Word Embedding." *Journal of Informetrics* 12(4):1099-117. doi: https://doi.org/10.1016/j.joi.2018.09.004.

Appendix A

TableA.1 Steps of TEI generation

Step		Proposition	Mathematical notation
I	Term Persistence	A term must appear in at least three time periods (years).	$\sum_{T} x_{it} \geq 3,\ x_{it} = \begin{cases} 1 & \text{if term i appear in time period t} \\ 0 & \text{otherwise} \end{cases}$
		A term must appear in at least seven records in the active period.	$\sum_{t=4}^{10} n_{it} \geq 7,$ n_{it} is the number of records contain term i in time period t
II	Novelty and Growth	The term cannot appear in as many as 15% of the base period records.	$\frac{\sum_{t=1}^{3} n_{it}}{\sum_{t=1}^{3} N_t} \leq 0.15,$ N_t is the total number of records in time period t
		The growth of the terms should be 1.5 times the growth of the total records.	$\frac{\sum_{t=4}^{10} n_{it}}{\sum_{t=1}^{3} n_{it}} \geq 1.5 \times \frac{\sum_{t=4}^{10} N_t}{\sum_{t=1}^{3} N_t}$
III	Community	Terms should be used by more than one organization.	$\sum_{J} y_{ji} \geq 2,\ y_{ji} = \begin{cases} 1 & \text{if organization j use term i} \\ 0 & \text{otherwise} \end{cases}$
IV	Scope	(Within-dataset filter) Terms should have an IDF-value higher than 1.	$\mathrm{IDF}_i \geq 1,$ IDF_i is the IDF value of term i in analysis dataset.
V	E-score Generation	An E-score for each term is calculated by the summary of two times the active trend, recent trend, and slope.	$\text{Active Trend}_i = \sum_{t=8}^{10} \frac{n_{it}}{\sqrt{N_t}} - \sum_{t=4}^{6} \frac{n_{it}}{\sqrt{N_t}}$ $\text{Recent Trend}_i = (\sum_{t=9}^{10} \frac{n_{it}}{\sqrt{N_t}} - \sum_{t=7}^{8} \frac{n_{it}}{\sqrt{N_t}}) \times 10$ $\text{Slope}_i = (\frac{n_{i10}}{\sqrt{N_{10}}} - \frac{n_{i7}}{\sqrt{N_7}}) \div 3 \times 10$ $\text{Escore}_i = 2 \times \text{Active Trend}_i + \text{Recent Trend}_i + \text{Slope}_i$
VI	TEI Generation	Terms with an E-score of no less than 1.77 are selected as emerging terms.	$\text{Escore}_i \geq 1.77$

Appendix B: Clustering with different k values.

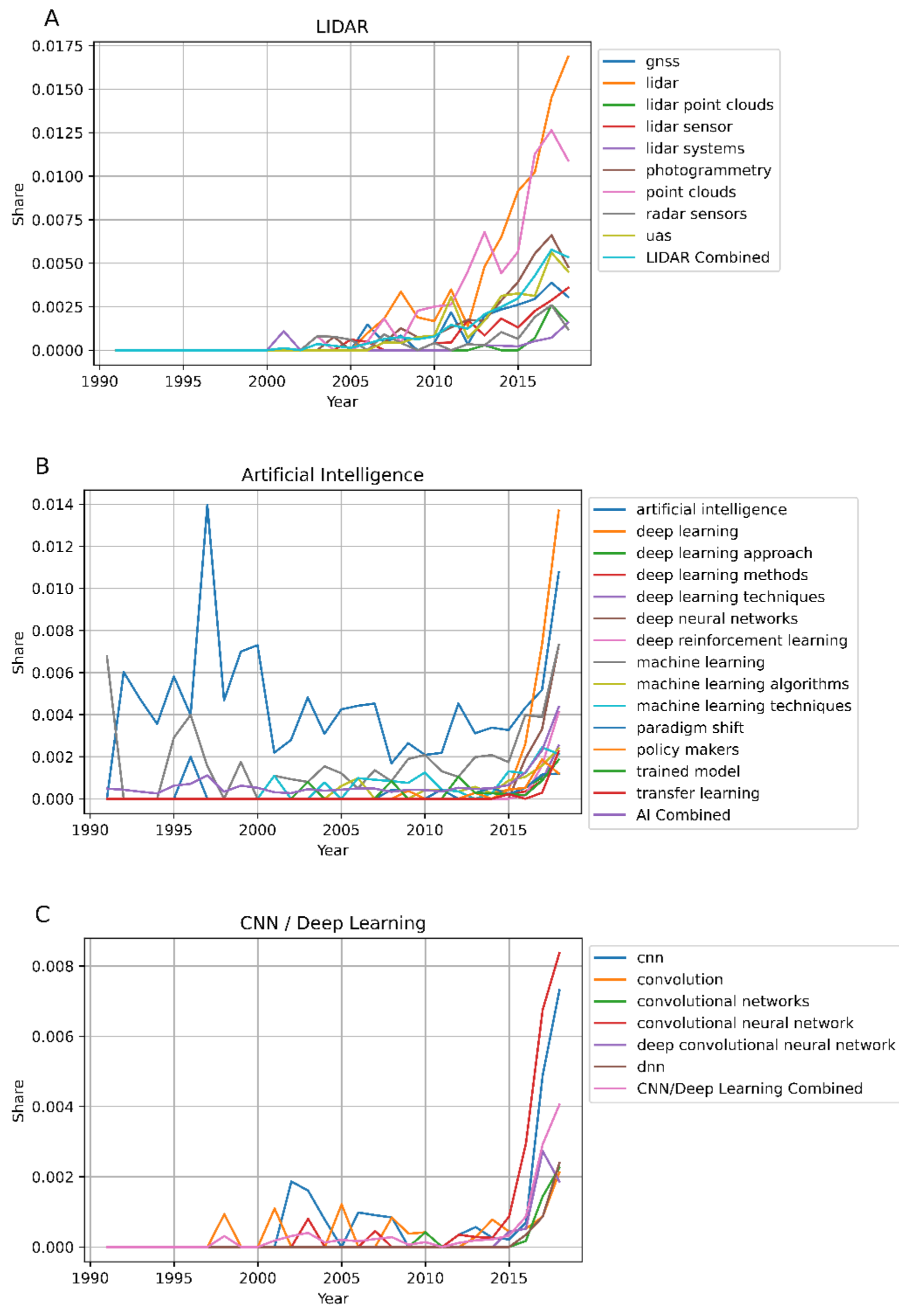

Figure B.1 Relative Growth in LIDAR (A), AI (B), and CNN/Deep Learning (C) topics. The y-axis is the share of publications in a given year to which an emerging term appears. Emerging terms shown in figures are clustered using the k-mean clustering algorithm with k=40.

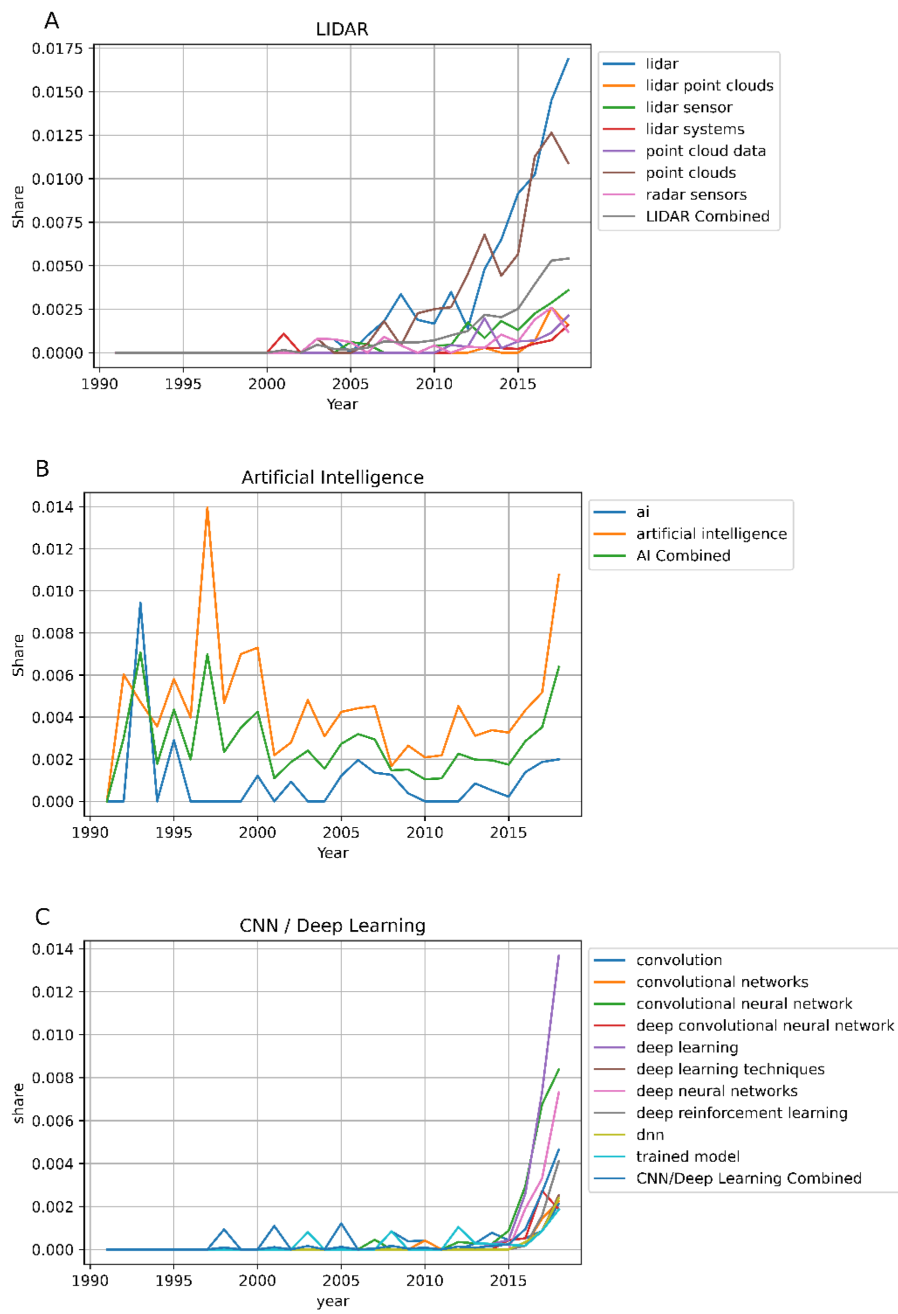

Figure B.2 Relative Growth in LIDAR (A), AI (B), and CNN/Deep Learning (C) topics. The y-axis is the share of publications in a given year to which an emerging term appears. Emerging terms shown in figures are clustered using the k-mean clustering algorithm with k=50.

Appendix C: Clustering using frequency weighted embedded vectors

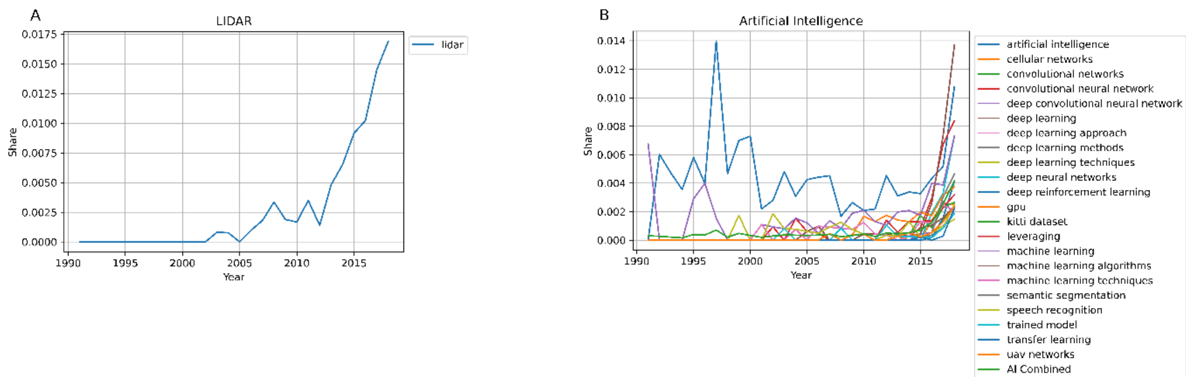

Figure C.1 Relative Growth in LIDAR (A) and AI (B). The y-axis is the share of publications in a given year to which an emerging term appears. For all noun-phrase terms, embedded vector for each emerging terms are combined using relative frequency of terms instead of its TF-IDF weight. Emerging terms shown in figures are clustered using the k-mean clustering algorithm with k=30.

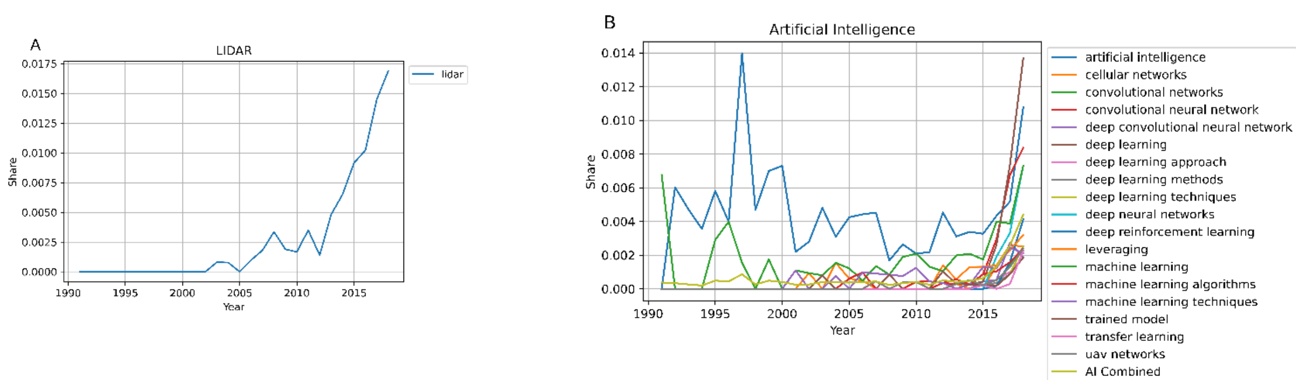

Figure C.2 Relative Growth in LIDAR (A) and AI (B). The y-axis is the share of publications in a given year to which an emerging term appears. For all noun-phrase terms, embedded vector for each emerging terms are combined using relative frequency of terms instead of its TF-IDF weight. Emerging terms shown in figures are clustered using the k-mean clustering algorithm with k=40.

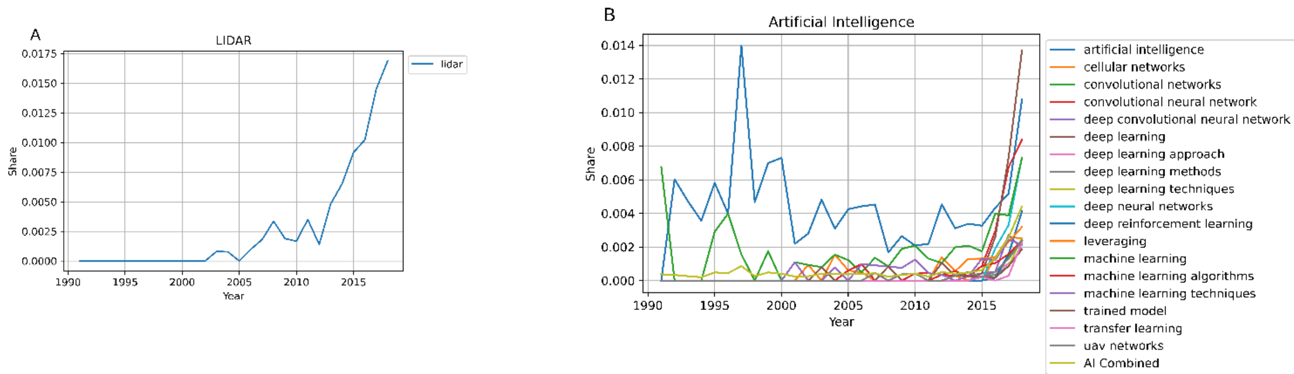

Figure C.3 Relative Growth in LIDAR (A) and AI (B). The y-axis is the share of publications in a given year to which an emerging term appears. For all noun-phrase terms, embedded vector for each emerging terms are combined using relative frequency of terms instead of its TF-IDF weight. Emerging terms shown in figures are clustered using the k-mean clustering algorithm with k=50.