Unraveling the capabilities that enable digital transformation: A data-driven methodology and the case of artificial intelligence

Mengjia Wu[1], Dilek Cetindamar Kozanoglu[2], Chao Min[3], Yi Zhang[1, *]

[1]Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

[2]School of Information, Systems, and Modelling, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

[3]School of Information Management, Nanjing University, China

Abstract

Digital transformation (DT) is prevalent in businesses today. However, current studies to guide DT are mostly qualitative, resulting in a strong call for quantitative evidence of exactly what DT is and the capabilities needed to enable it successfully. With the aim of filling the gaps, this paper presents a novel bibliometric framework that unearths clues from scientific articles and patents. The framework incorporates the scientific evolutionary pathways and hierarchical topic tree to quantitatively identify the DT research topics' evolutionary patterns and hierarchies at play in DT research. Our results include a comprehensive definition of DT from the perspective of bibliometrics and a systematic categorization of the capabilities required to enable DT, distilled from over 10,179 academic papers on DT. To further yield practical insights on technological capabilities, the paper also includes a case study of 9,454 patents focusing on one of the emerging technologies - artificial intelligence (AI). We summarized the outcomes with a four-level AI capabilities model. The paper ends with a discussion on its contributions: presenting a quantitative account of the DT research, introducing a process based understanding of DT, offering a list of major capabilities enabling DT, and drawing the attention of managers to be aware of capabilities needed when undertaking their DT journey.

Keywords

digital transformation; digital capabilities; bibliometrics; topic analysis; artificial intelligence

1. Introduction

Digital technologies such as artificial intelligence (AI) trigger significant changes to the organization's properties, widely known as digital transformation (DT) [1]. That is why DT has become an emerging phenomenon in strategic information research and industrial business practice [9, 29, 31]. At a macro level, all of society is experiencing profound changes as a result of the explosion in digital technology across various industries such as manufacturing, newspaper, and healthcare [85, 108, 111, 112]. At a micro level, organizations build their capabilities and take advantage of new digital technologies to realize innovations and create business value. The paper aims to shed light on our understanding of DT taking place at companies' micro level.

Yet, despite the vast interest in DT, there are still some significant research gaps in this domain. First, most conceptual definitions of DT are based on qualitative analyses, such as expert judgments or a literature review [1, 2]. An obvious drawback of qualitative methods is that such approaches may be biased due to experts' subjective opinions and limited cognitive [3]. Data-driven quantitative analysis, in contrast, yields objective insights based on the accumulated data and facts. However, quantitative approaches have rarely been used to characterize DT or the capabilities that enable DT at the current stage. This was one of Vial's most urgent and vital calls to researchers [1]. Second, many studies point out the significant role of capabilities for an organization's development [4-9]. However, almost no studies bring them together to understand their role in enabling DT. Lastly, when it comes to technological capabilities, most studies on DT are general and do not consist of any capability specific to a technology domain [4, 5]. They talk non-specifically about the realm of digital technologies instead of the technologies themselves, which makes it difficult to generate specific theoretical and practical implications from the results.

Hence, this paper aims to fill the three critical gaps: the lack of quantitative analysis on a definition of DT, the missing link between DT and the capabilities that enable it, and the lack of attention to understanding the capabilities needed to leverage specific digital technologies. By doing so, we hope to shed light on uncovering what entails DT when companies adopt digital technologies and what capabilities could help companies to manage this transformative process [9, 29].

To fill the first two gaps, this paper presents an empirical study to refine the definition of DT and reveal the digital capabilities that enable DT, treating digital technologies as a macro category. Then, it addresses the third gap by a case study focusing on one digital technology, artificial intelligence (AI), to exemplify the specific capabilities needed to leverage AI in a DT journey successfully. AI refers to technologies that interpret external data correctly, learn from such data as humans do - referring to the five attributes of emerging technologies summarized by Rotolo et al. [10], the emergence of AI has been well examined by both bibliometrics and technology management communities [11, 12]. Being a general-purpose technology, AI is listed as one of the seven digital technologies (including AI; autonomous vehicles; big data analytics and cloud

computing; 3D printing; IoT and connected devices; robots and drones; and social media and platforms) with the highest impact on the transformation of various industries [13]. Further, many studies in management literature emphasize the revolutionary characteristics of AI applications in terms of DT [14]. Hence, this paper chose it as one technology to delve into a detailed analysis to illustrate how to address the third gap in the literature.

More specifically, the overall goal of this paper could be concluded as answering the three following research questions:

- Question 1 (Q1): What is the definition of DT from a bibliometric perspective?
- Question 2 (Q2): What are the capabilities enabling DT?
- Question 3 (Q3): What are the AI capabilities enabling DT?

Bibliometrics is a set of statistical approaches that recognize the patterns in scientific activities by analyzing scientific documents, such as books, research articles, patents, etc. It has been used widely in science, technology, and innovation (ST&I) studies to identify research trends [15], predict emerging research topics [16], and discover explicit/implicit knowledge [17]. Unlike qualitative methods, bibliometric approaches rely on bibliographic indicators, such as keyword frequency, geographic distribution, and citation statistics, to generate derivate indices to quantitatively and measure the scientific behavior of research entities (e.g., researchers, institutions, and countries) among scientific disciplines and technological areas. Examples include scientific collaboration [18], technological convergence [19], changes in research focus [20], etc. Those derivate indicators provide solid quantitative evidence hidden behind the scientific publications, which is sorely needed in DT studies.

This study is to leverage bibliometrics to seek quantitative evidence of exactly what DT is and the capabilities needed to enable it successfully. The framework devised for this study integrates a topic tracking method called scientific evolutionary pathways (SEP) [21] with a novel method of identifying topic hierarchies named hierarchical topic tree (HTT). Specifically, the SEP is to track the change of research foci on the DT research in the past decade and provide clues on locating certain most recent interests of the community. The HTT, complementarily, profiles the research landscape of the DT research into a hierarchical structure and helps identify the quantitative composition of core research topics in the DT research. Considering DT is a rapidly developing research area, revealing its scientific evolution and clarifying its composition will offer insights into its definition and enabling capabilities in dynamical and quantitative manners. By incorporating these two methods with network analytics and a literature review, each of our three research questions can be answered empirically rather than subjectively.

More specifically, we collected 10,179 scientific articles from the Web of Science and 9,454 patents related to AI from the Derwent Patent Citation Index. Through SEP analysis on the collected dataset of scientific papers, we identified the evolutionary patterns of research topics in the DT literature, enabling us to distill a general definition of DT (Q1). To unravel the specific capabilities enabling DT (Q2), we applied HTT coupled with a literature review of specific

papers on digital capabilities to arrive at a comprehensive categorization of the resources and competencies needed to successfully undertake a DT journey, with the specific capabilities from 39 core papers classified as dynamic capabilities, technological capabilities, platform capabilities, and other capabilities. Lastly, taking AI as our focus technology, we applied an HTT analysis to a corpus of patents on AI and conceptualized a four-level model to guide AI-enabled DT (Q3). From the bottom to the top, the model progressively presents data collection & transmission capabilities, bridging capabilities, algorithm capabilities, and application capabilities required for companies to leverage AI in their digital transformation process.

The contributions of this work are many-fold. Through the SEP and HTT analyses, our research accumulates solid quantitative evidence to define DT. This evidence will enable researchers to deepen and broaden definitions of DT in future studies. It may also serve as a pointer to emerging topics in DT research and its evolution. The detailed analyses of bibliometrics and patent database facilitate to determine the diverse capabilities required for DT generally and the specific capabilities required for AI. Further, our work yields practical implications about capabilities that managers could tap into while dealing with the challenges of DT. These are the core elements companies might use to ensure their DT journey is a successful one.

The rest of this paper is organized as follows. In Section 2, we briefly review relevant work on DT, digital capabilities, and bibliometric approaches. Section 3 presents the details of our collected datasets and methodology. The empirical study of unraveling DT enabling capabilities follows in Section 4. Section 5 presents the case study on AI, and Section 6 summarizes the study with a discussion, our limitations, and future directions of research.

2. Related work

2.1. Digital transformation and capabilities

Digital technologies are many and varied, ranging from AI and blockchain to 3D printing and big data analytics. However, despite their diversity, they all share a few key features that make them general-purpose technologies [14, 22-24]: re-programmability, homogenized data, ubiquity, a self-referential nature, a layered architecture, and automation of data generation through many sources, such as sensors and machine learning algorithms.

The characteristics of digital technologies supply three critical capacities to firms [23]: openness, affordance, and generativity. Openness refers to open innovation practices. Affordance points out possibilities or opportunities for action, and generativity concerns the capacity exhibited by digital technologies to produce change through various uncoordinated actors/entities. Thus, digital technologies have a great potential to impact many industries and companies.

Due to the rapid diffusion of digital technologies, organizations are under constant pressure to transform their business models, products, services, and processes. Even though a new technology means transformation, studies on digital technologies agree that the digital era has,

beyond compare, resulted in technology-enabled transformation [13, 14, 23]. However, although the term DT is widely used in the literature, it is used without a common acceptance of its meaning, as clearly indicated in many studies [1, 25]. That is why this study analyzes the concept and its evolution, as revealed in Subsections 4.1 and 4.2.

The uniqueness of technology-enabled transformations in the digital era comes from two developments. The first is the switch from one-off transformations to continuous ones. The second is the expansion from individual company transformations into complex multiple organization transformations [26]. Continuous DT underlines recurring activities around digitization and the digitalization of products, services, and processes. Digitization describes the conversion of analog to digital information and processes in a technical sense. Digitalization applies digitizing techniques to provide the affordances of digital technologies [24]. DT has resulted in fundamental business models that no one-off transformation can achieve [27].

Multiple organizational transformations are a fact of the digital era [28]. Digital ecosystems/platforms are radically changing traditional value chains and conventional industry structures. A digital platform is a combination of hardware and software that provides standards, interfaces, and rules that allow the providers to add value and interact with each other and/or users [27]. For example, the music sector transformed into a new digital ecosystem where computers, smartphones, social media platforms, and software have connected and transformed many organizations simultaneously [28]. As another example, Amazon's Kindle brings together a complex ecosystem of companies from the computer industry, consumer electronics, Internet search, online retailing, book retailing, telecommunications, and publishing [22].

Recent studies have sought to understand how to manage the continuous DT journeys of companies across different sectors [1, 25]. That is why many companies consider that building a set of dynamic capabilities is a strategic imperative for ensuring survival in the digital age [2, 9].

DTs highlights the necessity for building dynamic capabilities in order to sense its needs, opportunities, and threats, mobilize its resources and transform operations to create new services, build new products, or come up with new business models and processes [27]. However, considering that DT is a continuous process and that it draws on digital platforms, it is necessary to identify all the capabilities critical to this process. Only then can companies get ready for the future by building assets and competencies that align their business strategies with their potential capabilities and using digital platforms to succeed in the transformation [2].

However, the task of unlocking the capabilities needed for DT is not straightforward. The most influential theory on the capabilities needed at the organization level comes from dynamic capabilities theory. Dynamic capabilities refer to a firm's capacity to "integrate, build, and reconfigure internal and external competencies to address rapidly changing environments" [29]. Dynamic capabilities are considered higher-level capabilities that: (a) enable organizational learning; (b) create new combinations of assets; and (c) renew operational (or ordinary) capabilities [30]. This theoretical paradigm puts forward three dimensions of dynamic

capabilities: sensing, seizing, and reconfiguring [29]. The traditional model is very simplistic. However, a study of around 2,500 articles on dynamic capabilities in the strategy field found many and diverse theoretical approaches to dynamic capabilities [31].

Yet focusing on dynamic capabilities alone is highly limiting, especially considering that DT is a process where digital technologies play a crucial role. Technological capabilities represent a firm's knowledge base and intellectual capital to ensure innovation [32, 33]. The literature is full of studies assessing technological capability through a firm's patents since they represent knowledge the firm created [34, 35].

By responding to the calls in the literature [1], we aim to address the knowledge gap in the literature regarding the fuzziness of the DT concept (Q1) and by identifying the capabilities that can assist DT (Q2). To do so, we conduct a literature review and a bibliometric analysis with a set of unique approaches, as described in Section 3.

2.2. Bibliometrics and topic analysis

Bibliometrics is the discipline of tracing scientific activities by quantitatively analyzing scientific literature and documents [36]. The incorporation of bibliometrics with emerging information technologies like AI (text mining and natural language processing, especially) and advanced computational models has introduced a new pathway for bibliometric studies [17, 37]. We call it *intelligent bibliometrics* [38]. Our pilot studies have demonstrated the effectiveness of these cross-disciplinary incorporations [17, 21, 39].

Topic analysis, a content-based analysis method in bibliometrics, extracts research topics from a collection of scientific documents and analyzes the relationships between them. In most cases, topic extraction is conducted via clustering or a classification algorithm based on term co-occurrence or the citation relationships between documents [40-42]. The extracted topics are normally represented as sets of terms or phrases partitioned into research disciplines [43], technological areas [19], or other latent relationships [44]. Further analysis of those topics can help clarify cross-/inter-/multi-disciplinary interactions or predict future emerging research topics/interests [20, 45].

The traditional topic analysis approaches extract topics in a flat structure and analyze them solely from a lateral perspective [40]. This method neglects two crucial characteristics of research topics: how the focus of research evolves over time and the research attention given to the topics, i.e., the topic hierarchy. First, the topics of focus in scientific publications are not static; they are constantly changing and evolving over time [21]. These shifts in focus may be the result of multi-disciplinary interactions, disruptive factors, or ground-breaking findings [46]. Capturing these trends reveals the evolutionary patterns in research attention from a dynamic standpoint. Second, hierarchy is another innate structure rooted in scientific knowledge composition. Hierarchies profile research topics in a top-to-bottom structure, with different topic granularities and subordinate relationships attributed to topics on different levels [47, 48]. Combined, these

two characteristics of research topics can mean topic analysis provides far greater and deeper insights.

DT is a newly emerging research topic that has experienced rapid change and development over the past decade [1, 49]. Hence, revealing the evolutionary patterns and hierarchies of DT research topics is an important part of 1) developing a dynamic understanding of the definition of DT and 2) identifying the key enablers of DT. Driven by these expectations, our methodological framework incorporates a SEP analysis [21] to reveal the evolving topics of interest in DT studies and an HTT analysis that decomposes the knowledge structures in DT research from a hierarchical perspective.

3. Data and Methodology

This section begins with details of our case data, followed by a detailed description of our analysis framework.

3.1. Data collection

The answers to Q1 and Q2 can be found from a quantitative analysis of academic articles as such studies majorly focus on academia's attitudes towards DT and its implications among companies, countries and society. On the other hand, answering Q3 requires empirical evidence from industry practice, for measuring which we selected patent data that covers a broad range of specific AI techniques and their applications in real-world DT cases. On the other hand, patent data covers a broad range of specific AI techniques and their applications in real-world DT cases, is an appropriate source to identify specific AI technologies for enabling DT.

The Web of Science (WoS), owned by Clarivate Analytics, is a multidisciplinary scholarly database with 74.8 million scientific publications from over 21,100 journals. Despite its slightly narrow coverage, the WoS provides more well-curated bibliographic metadata than Google Scholar [50-52] and concentrates on journals with higher impact than Scopus [53, 54]. Besides, it provides download access to batch textual abstract data for large-scale data analysis. For patent collection, the Derwent Patent Citation Index (DPCI) is the most comprehensive patent database that contains 39 million patent citations covering all technological sectors [55]. From the aforementioned considerations, we respectively selected the WoS and DPCI as our data sources for collecting scholarly documents and patents.

The following search strategy returned 10,179 articles related to DT from WoS:

TS = ("digit transfor*" OR "digitization*" OR "digitisation*" OR "digitalization*" OR "digitalisation*" OR "digit* capabilit*" OR "digit* platform*" OR "digit* tech*" OR "digit* innova*" OR "digit* competence*" OR "digit* mind*" OR "digit* activit*" OR "digit* practice*" OR "digit* manag*") AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Timespan=2010-2020, search date:24 September 2020*

These papers constitute Dataset 1. From these data, we additionally prepared a second, more focused dataset of 913 papers published in journals our experts deemed to be of "high-quality" as dataset 2, which have a stronger theoretical basis for management and are more insightful in revealing DT capabilities. The list of the journal titles is provided in Appendix A. Dataset 3 was assembled to support the case study. It comprises 9,454 patents relating to AI drawn from the DPCI. Given the general concept of AI covers such a broad range of areas, we narrowed our search to only patent titles that contained "artificial intelligen*") with the following search strategy:

TI = ("artificial intelligen") AND IP=(G06* OR H04* OR H01* OR G11* OR G10* OR G01* OR G02* OR H05* OR H02* OR H03* OR G09* OR G05* OR A63* OR G08* OR G03* OR B60* OR G07* OR F24* OR A61* OR B65* OR B23* OR B81* OR B25* OR C08* OR A45* OR B01* OR C25* OR C09* OR B64* OR C23* OR F16* OR A44* OR C12* OR B32* OR C03* OR B62* OR F04* OR B29* OR B41* OR B24* OR F25* OR F28* OR E04* OR F21* OR G12* OR G04* OR G16* OR C01* OR B66* OR C07* OR B22* OR A47* OR A01* OR B82* OR B05* OR C22*), search date:06 November 2020*

In summary, the three datasets are:

- Dataset 1: 10,179 scientific papers related to DT retrieved from the WoS.
- Dataset 2: a subset of Dataset 1 comprising 913 articles published in high-quality journals.
- Dataset 3: 9,454 AI patents from the DPCI.

3.2. Methodology

The framework of our methodology is illustrated in Figure 1. It involves four main analysis techniques: SEP, HTT, network analytics, and literature review. SEP charts the course of innovation from one research topic to the next; we applied this method on dataset 1 to capture the dynamic changes of DT definition in the latest decade. HTT decomposes the technologies into vertical corridors of research and development; we applied HTT on dataset 2 to decompose the knowledge into a hierarchical structure and identify the capabilities of DT. Further, HTT was also applied to dataset 3 to profile a hierarchical AI technique landscape in DT. At the same time, network analytics and literature review thoroughly run on the results of SEP and HTT to reveal the topics' grouping characteristics and yield an in-depth analysis of data related to specific topics.
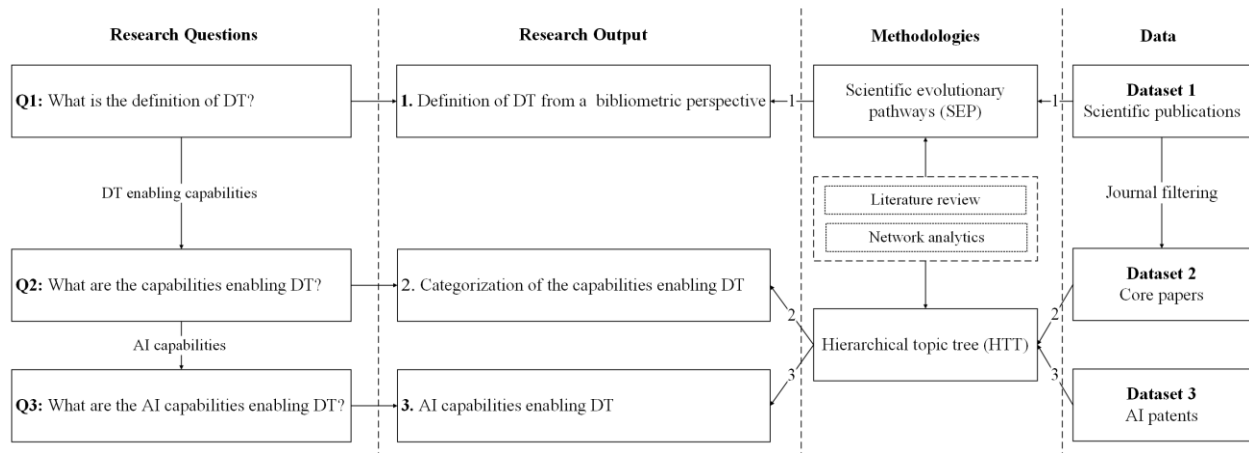
Figure 1. The research framework

3.2.1 Data pre-processing

Before using the data for the SEP and HTT analyses, we applied VantagePoint's [1] natural language processing (NLP) function to convert the datasets into a dictionary of raw words and phrases. We then executed a term clumping process [56] that removes noise and consolidates synonyms to arrive at a final list of topic terms. From this list, we selected the 5000 terms with the highest frequency from Datasets 1 and 2, and terms with a frequency of greater than one from Dataset 3 for further analysis. The stepwise results are given in Table 1.

Table 1. Stepwise results of the term clumping process

| Step | Description | #Terms | | |
		Dataset 1	Dataset 2	Dataset 3
1	Raw terms retrieved with NLP	253,162	24,203	161,722
2	Consolidated terms with the same stem, e.g., "information system" and "information systems"	220,812	20,530	154,575
3	Removed spelling variations, removed terms starting/ending with non-alphabetic characters, e.g., "Step 1" or "1.5 m/s", removed meaningless terms, e.g., pronouns, prepositions, and conjunctions	199,410	18,398	146,756
4	Removed general single-word terms, e.g., "information" *	174,880	15,281	132,655
5	Filtered technological terms suggested by experts	-	-	10,279
6	Consolidated synonyms based on expert knowledge, e.g., "co-word analysis" and "word co-occurrence analysis"	164,433	14,918	849
7	Eliminated all but the top 5000 most frequently occurring terms	5,000	5,000	-
8	Eliminated all terms occurring only once	-	-	226

*Note: Considering the most concepts of single-word terms are enriched by multi-word terms from different perspectives, e.g., "information" vs. "information system", we decided to remove those general single-word terms. However, some multi-word terms may be consolidated into their related non-general single-word terms in Step 2 (like "classification" and "classification method"), those non-general single-word terms were preserved.

3.2.2 Scientific evolutionary pathways

Referring to an assumption that scientific novelty usually recombines established knowledge [57, 58] and scientific evolution over time, SEP describe scientific evolution by tracing the change of research topics from a corpus of scientific documents published in a given time period [21]. We ran this SEP algorithm on the pre-processed documents in Dataset 1, and yielded topic evolutionary pathways at a macro level. The specific procedure is described as follows:

Step 1: All documents were assigned into batches according to their publication year. Given the entire vocabulary of the dataset as a feature space, each document was represented by a term vector, filled with the co-occurrence between this document and a given term.

Step 2: In the first batch (Batch 0), we grouped all documents into one initial topic and considered it as the starting point of this evolutionary pathways – with an assumption that the origin of a concept, a research domain, or a technique might be rough, and its evolution is the establishment of interactive knowledge clusters but with boundaries. An empty list of current topics was created, and we added the initial topic to the list.

Step 3: Starting from Batch 1, we measured the similarity between each document in this batch and all current topics to determine which topic this document belongs to and to which extent this document shares similar content with this belonged topic. Particularly, we grouped documents that belong to but are slightly different with a current topic into new topics, and the relationships between the current topic and these new topics were identified as predecessor-descendant relationships.

Step 4: We iterated Step 3 for each batch until the end of the dataset.

The output of the SEP algorithm includes a list of topics and their predecessor-descendent relationships. We visualized the evolutionary pathways in the form of science maps, with the aid of Gephi [59] – each topic is represented as a node and their predecessor-descendant relationships are described as a direct edge (from predecessors to descendants). An algorithm of community detection in Gephi, called modularity, was used to group proximate nodes into communities, indicating relevant research contents and technologies and their components, and nodes in the same community are annotated with the same color.

3.2.3 Hierarchical topic tree

The HTT algorithm aims at uncovering the hierarchies of research topics. It works on a co-occurrence network of topic terms or technologies, building hierarchies based on two

assumptions: 1) that the connection from a subordinate to its parent is stronger than to its neighbors, and 2) that parents receive more attention than their subordinates. The pseudocode for the algorithm is given in Algorithm 1, and a summary of the steps follows:

Step 1: Construct a weighted co-occurrence network (graph) $G = (V, E)$ as follows:

$$E_{V_i V_j (i \neq j)} = \begin{cases} CF(V_i, V_j) & if\ V_i\ and\ V_j\ co-occur\ in\ at\ least\ one\ document \\ 0 & else \end{cases}$$

where V is the set of nodes representing the topic terms, E is the set of edges representing the co-occurrences of connected topic terms, V_i and V_j are topic terms, and $CF(V_i, V_j)$ is the co-occurrence frequency, i.e., the number of documents in which both terms appear.

Step 2: Apply the maximum spanning tree algorithm [60] to construct an undirected maximum tree structure from the graph, this tree layout generated from the network will be the prototype of our final HTT.

Step 3: Assign directions to the edges generated in Step 2. Directions are set according to the following criteria: a) every document has a unit of resource and allocate it evenly to the terms it has, resulting in each node representing one unique term has a resource allocation aggregation from all the documents; b) the root node has the most resources; c) superior topics have more resources than their subordinates; and d) directions move from greater to fewer resources. The governing equation is:

$$R_{V_i} = \sum_{m \in D^{V_i}} \frac{1}{T(D_m^{V_i})}$$

where R_{V_i} is the resources of V_i, D^{V_i} is the subset of documents that containing the topic term V_i, and $T(D_m^{V_i})$ is the topic term count of the mth document in D^{V_i}.

Step 4: Group topologically similar leaf topics and identify the leaf clusters under a mutual parent node in the tree. With the edge directions, edge weights and clustering information assigned to the prototype in Step 2, we will obtain a finalized HTT.

Algorithm 1. Pseudocode of the hierarchical clustering algorithm

---

Given:
A set T of topic terms under the same root node $\{t_1, t_2, \dots t_k\}$
A function that gets the edge weight of t_i and t_j in G: $weight(t_i, t_j)$

# return one cluster if the number of leaf nodes is less than 4
if $k < 4$:
  $C = \{t_1, t_2, \dots t_k\}$

```
else:
    for i = 1 to k
    C_i = {t_i}
    end for

# A function that returns the average edge weight of avg_weight(C_m, C_n)
avg_weight(C_m, C_n):
    return average(weight(t_i, t_j))

for t_i in C_m for t_j in C_n
    C = {C_1, ..., C_n}
    while C ∩ T ≠ ∅ do
    (C_max1, C_max2) = avg_weight(C_i,C_j) for all C_m, C_n in C
    remove C_max1 and C_max2 from C
        add C_max1 ∪ C_max2 to C
end while

Output: the final clustering result C
```

We ran the HTT algorithm on Datasets 2 and 3 to produce two different trees: one for DT and the other for AI technologies. As a backbone schematic of research, the HTT from Dataset 2 (high-quality academic papers) was used to help arrive at a standard definition of DT and to build categories of the capabilities needed to enable DT. The HTT from Dataset 3 was used in our case study to help identify which AI technologies are typically adopted in a DT process.

3.2.4 Expert knowledge via a literature review

Based on the SEP from Dataset 1 and the HTT from Dataset 2 generated, we could retrieve a set of papers talking about digital capabilities for our special interests. We then manually read the papers and selected those focusing on digital capabilities from a company level, further narrowing down the review scope. The other papers discussing digital capabilities from individual-, society- or other levels are discarded. Two authors conducted a literature review on the filtered papers and summarized a categorization of digital capabilities. The concrete review protocol and the process will be presented in Subsection 4.3, along with the reviewing results.

4. Empirical study: Unravelling capabilities enabling DT

4.1. SEP analysis

The SEP analysis helped us to define DT from a bibliometric perspective in answer to RQ1. The scientific evolution map for DT research is shown in Figure 2. It traces the changing focus of academic research over the last decade. Each node represents a topic, and each edge indicates the predecessor-descendent relationship between its connected topics. The colors indicate the four

topic communities detected by Gephi. Particularly, in a SEP map, the evolution is visualized through two aspects: (1) a pair of square brackets in each topic label indicates the year when this topic was born, and (2) a direct edge connects a predecessor and a descendent, and thus a pathway consisting of a set of connected topics may describe how topics in a research area evolve over time.

Figure 2. The SEP of the DT research between 2010 and 2020

Observing the time labels of those topics, we summarized three stages of DT research in the last decade. From 2010 to 2012, discussions in this field mostly focused on fundamental DT concepts, including digital technologies, digitization and information communication technologies [22, 75, 94]. Such general topics indicated DT research in this period is at an infancy stage of clarifying the concepts and basic elements that constitute DT. Since 2013, the implementation of DT in industry field had rapidly emerged as the key focus of DT studies, especially in the manufacturing, educational and healthcare sectors [such as 108, 109, 111, 112, 118, 133, 140]. In the same period, digital platforms were also brought out for their values in accoupling with

digital technologies to enable DT [61, 62, 63, 103, 114]. For this stage, we concluded that the impact of DT in real-world practice and DT strategies were the new interests of researchers. The last stage is 2018-2020, in which period the emerging topics were advanced technologies like remote sensing, social media, blockchain, artificial intelligence, immerse reality, etc [65, 66, 68, 103, 136, 139]. This stage reflects the most recent trend of DT development of constantly absorbing new technologies and may trigger industry revolution in the future.

Apart from the dynamic characteristics, research topics also present clustering patterns along with the evolutionary pathways. Generally, four topic communities were detected in the SEP: Green (#I) encompasses the fundamental concepts of DT research at a macro level. Orange (#II) indicates the initial development of digital technologies researched in DT and highlights the heavy involvement of interactive technologies. Blue (#III) uncovers communications technology (CT) and other CT-based emerging technologies that are prominently related to DT, like mobile CTs and AI. Lastly, pink (#IV) denotes digitization processes and the transition of DT from theory into practice.

#I Digitization (green): This community can be thought of as a birds-eye view of the spectrum of research into DT. Many of the major evolutions emanate from the industries that need or benefit from DT, such as healthcare, education, and manufacturing *(relevant topics: healthcare [2017], COVID-19 [2020], teachers [2013], and manufacturers [2014])*. The milestones along the major pathways include: changes to the fabric of industry itself *(Industry 4.0 [2018], industrial internet [2020])*; digitizing objects *(digital text [2015], digital media [2018], digital images [2016], 3D digitization [2020])*; and research methods *(questionnaires [2019], interviews [2019], web-based surveys [2020], semi-structured interviews [2020])*. Of the four communities, this one has the widest scope.

#II Digital technologies (orange): Derived from Community #I, this community reveals the first offshoots of DT – the technologies developed. The topics include digital platforms, social networks, advanced interaction media, and immersive reality [61-63], which are mostly interactive and user-engaged. This community also acts as a bridge to Community #3, where information communication technologies (ICTs) emerge.

#III ICTs (blue): Community #III encompasses *smartphones [2016], mobile technologies [2017]* and *mobile devices [2018],* highlighting the developing trends in mobile communications [64]. It also includes the emerging topics of data analytics, artificial intelligence, and sustainability *(artificial intelligence [2018], principal component analysis [2019], and logistics models [2020], sustainable development [2019], and smart cities [2020])* [65-67].

#IV Digitization processes (pink): While Community #I shows the theoretical beginnings of DT, Community #IV shows the practical outcomes. Many of the topics in this community are either key enablers to DT *(big data [2017], IoT [2018], remote sensing [2018], Blockchain [2019], etc.)* or digitization solutions *(technology integration [2015], the sharing economy [2018], manual digitization [2019])*, which provides guidance for companies to realize a successful DT [68, 69].

Overall, this SEP analysis reveals how the research into DT has evolved over time. The four different topic communities reveal the "entity" (#I), the "technologies" (#II and #III) and the "significant change" that has occurred as a result (#IV), providing quantitative evidence of how the theoretical foundations of DT can become a reality. These four communities accord with Vial's [1] definition that "DT is a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies".

4.2. HTT analysis

Figure 3 shows the HTT for Dataset 2. The weight of edges represents the normalized value of the co-occurrences between connected items - i.e., how many publications these two items appear together in our dataset 2. Aiming to remove trivial branches and ensure a concise final tree, we usually retain no more than four layers in the tree and in this case, we set a threshold for the co-occurrent frequency as 8. The finalized result was a tree containing 41 nodes with the weights of all edges within the range of 0.1-1.
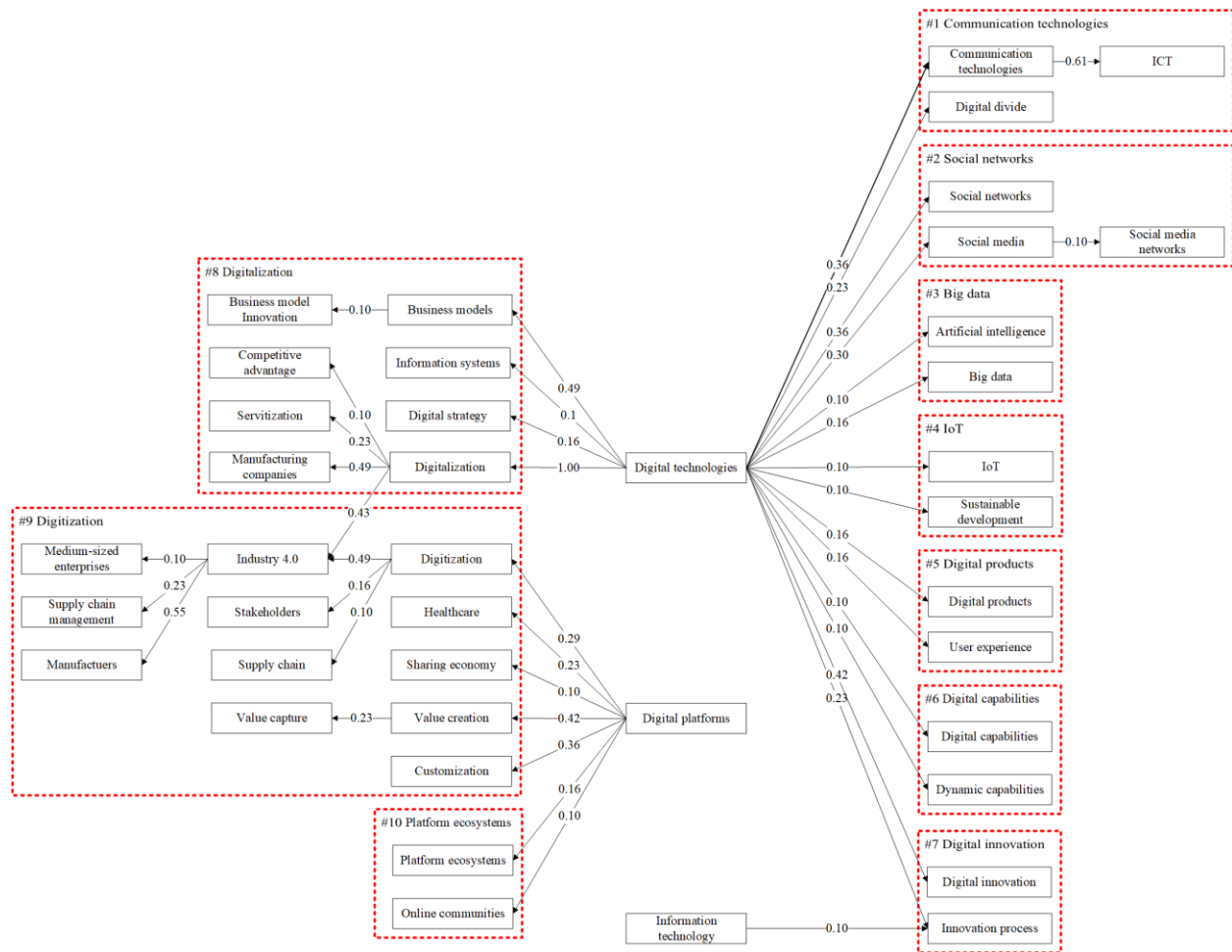

Figure 3. The HTT discovered in DT research

HTT reveals topic relationships from a cross-sectional and vertical perspective. Compared to Figure 2, the HTT in Figure 3 yields some different insights into the state of digital technologies. For example, in Figure 2, *digital technologies* is not the starting node of the evolutionary pathway; however, it is an early milestone and acts as a bridge between inception and realization. Yet, in Figure 3, *digital technologies* is the most dominant node. Therefore, both figures indicate that the development of digital technology is a critical enabler of DT. *Digital platforms* and *information technology* are the other two nodes in the tree without parent nodes, which means these two topics are equivalently important with digital technologies to DT. The remaining nodes are divided into 10 topic clusters. We named each cluster after the node with the strongest connection to its parent. Although many of the topics in the HTT and evolution map are not the same, there are frequent correspondences between the two.

On the right side of the tree, there are seven topic clusters subordinate to the *digital technologies* root node. These seven clusters represent either specific digital technologies or the business implications of those technologies. The top four are *ICT* (#1), *social networks* (#2), *AI* (#3), and *IoT* (#4) – all digital technologies. From a full reading of some of the papers in these clusters, we find that disruptiveness is a characteristic shared by all topics [63, 69, 70]. By disruption, we mean they have the capability to shake up industries, trigger the development of new business models, and segment markets in new ways [71]. The papers in these clusters articulate how the technologies they discuss can enable DT [64, 66, 68, 69, 72-74], and they present empirical evidence to prove it [62, 65, 75]. Thus, the social implication of the technologies in this cluster is also a research focus. For instance, the explosion and imbalance of ICT development is claimed to be one of the causes of the digital divide – a prominent social problem in implementing digital technologies [76, 77], while IoT is recognized to have the potential to promote sustainable development by industry [65, 67].

The remaining three of the seven are *digital products* (#5), *digital capabilities* (#6), and *digital innovation* (#7). These reflect the business implications of digital technologies.

The articles pertaining to *digital products* #5 typically aim to promote the development or improvement of digital products based on digital technologies [78, 79]. Here, user experience is usually the key evaluation indicator [80]. The focus in *digital capabilities* #6 is on the capabilities required by organizations and individuals in the DT process, which are divided into digital [81, 82], dynamic [83-86], and their combination [31, 87]. A few studies have attempted to shed light on the connotation of some capabilities. However, there is still no guidance in the form of widely-accepted and comprehensive categories of the specific capabilities required for a successful DT. To further explore this question, we conducted a literature review, the results of which are summarized in Subsection 4.3. The articles grouped under *digital innovation* #7 discuss how to realize digital innovation by using digital technologies [88-90]. Here, *information technology* is also a parent of *innovation processes*. This indicates that information technology is something of a universal research topic but has a specific power to drive digital and/or business innovation [91, 92].

The last subordinate cluster of *digital technologies* is *digitalization* (#8) at the left top. The articles in this cluster discuss several prevalent issues for company management and industry governance in the DT process, such as alternative and innovative business models [93, 94], digital strategy scheduling [95, 96], the development and enhancement of business servitization [97, 98], and how to maintain one's competitive advantages through digitalization [99, 100]. Another notable highlight in this cluster is the frequent mention of manufacturing as a representative industry experiencing digitalization [101, 102].

The next parent, *digital platforms*, is linked to two subordinate clusters: *digitization* (#9) and *platform ecosystems* (#10). Papers in the *digitization* #9 cluster claim the significance of digital platforms in realizing business value [61, 103, 104] and accelerating the digitization process [85]. Further, this cluster contains several branches highlighting the highly-relevant sectors of digitization in practice, including supply chains [105-107], manufacturing [108-111], and healthcare [112, 113]. Papers relating to *platform ecosystems* #10 explore the establishment, development, and implications of platform-based digital ecosystems [114-116].

Intriguingly, our algorithm placed the topic *Industry 4.0* at the convergence of two parent nodes, *digitization* and *digitalization,* which means that, at the macro-level, realizing industry 4.0 requires both digital technologies and transformative business practice [117, 118].

In summary, this topic tree quantitatively reflects the composition of research topics in the field of DT, highlighting that technologies and platforms are two essential enablers of the transformation process. The four digital technologies most frequently studied to promote DT are *ICT*, *social networks*, *AI*, and *IoT*. Establishing digital platforms can empower DT by creating digital ecosystems and providing systematic business digitization approaches [35, 61]. Among all the clusters, *digital capabilities* #6 specifically gathers papers discussing capabilities that enable DT, which corresponds to our RQ2. Hence, in the next subsection, we review the papers in this cluster to unravel the specific capabilities needed to enable a successful DT.

4.3. Categorizing the capabilities that enable DT

The findings outlined in Subsections 4.1 and 4.2 highlight that no overall category has emerged from the exacted terms but rather only categories that encompass technologies and digitization. From Figure 2, we see digitization plays a substantial role. This finding confirms the emerging newness of DT and the lack of a coherent theory behind it. It also shows the importance of digitization as a broad category of observing transformations taking place through digital technologies. But, more importantly, our findings clearly highlight that DT is highly connected with digital technologies and digital platforms, as Figure 3 shows. Hence, we must revise Vial's definition of DT [1], proposing that DT be defined as a process that aims to improve an entity by triggering significant changes to its properties through combinations of digital capabilities, technologies, and platforms.

Given that DT is a process of change, it is critical to understand what capabilities can help its development and management. The *digitization* cluster in Figure 2 and the *digital capabilities* cluster #6 in Figure 3 point us to the academic work on this subject. *Digital capabilities* #6 represents 59 articles, which we read to determine whether they offer any clear and specific insights into the capabilities required for DT at a company level. Here, clear and specific means definitions are given and/or constructs have been developed and used to measure them (see Appendix B). Of the 59 papers, 31 satisfied these criteria and were included in the review. The other papers typically focused on society-level issues, such as measuring the Industry 4.0 readiness of manufacturing in the EU [119], or on individual-level topics, such how to assess digital skills in citizens [120]. These papers were discarded.

The majority (68%) of the reviewed papers are published in the period of 2018-2020. Empirical studies consist of 83% of articles and IT is the major technology is the main technology focus in these studies. One third of papers have no specific technology focus. Both authors read and classified the 31 articles based on the capabilities discussed, as listed in Table 2.

We find that the capabilities fall into four broad categories: dynamic, technological, platform, and other. More than half of the articles (55%) mention dynamic capabilities, consisting of 8 capabilities as given in Table 2. Only 16% of papers indicate technology capabilities, where five individual capabilities stand out. Another group (16% of papers) mention platform capability with two specific capabilities. The "Other category" comes from nine capabilities driven from 10 articles. While platform capabilities highly depend on digital technologies, the other category consists of quite broad ones with no clear link to digital technologies (such as knowledge management and R&D capabilities).

Dynamic capability is the larger and more diverse category, but mostly including traditional and non-traditional variations of three key concepts: sensing, seizing, and transforming [29]. One article, driven by a study of 208 innovations in the insurance industry, uses the term "transformative capabilities", referring to sector-specific capabilities, such as developing services that fulfill customer needs, exploiting data for risk assessment, and underwriting [121].

Table 2. The 31 articles of capabilities enabling DT

Capability category	Capabilities	Source
Dynamic capabilities	Sensing, seizing, transforming	[83, 84, 86, 121, 122]
	Digital sensing, digital seizing, digital transforming	[9]
	Absorptive capacity	[83]
	Integrative capabilities	[83, 104, 123]
	Relational capabilities	[83, 123, 124]
	Innovative capability	[100, 104]
	Dynamic managerial capabilities	[125, 126]
Technological capabilities	Digital capabilities	[31, 72, 81, 82, 127]

	Dynamic IT capabilities	[25]
	Big data capabilities	[128]
	Information analytics	[129]
	Relational and information processing capability	[130]
Platform capabilities	Platform capability	[25, 85, 124]
	Platform utilization capabilities	[125, 126]
Others	Business process management capabilities	[87, 131]
	Project capabilities	[132]
	Organizational learning capabilities	[133]
	Customer service capabilities	[134]
	R&D capabilities	[135]
	Production capabilities	[131, 135]
	Knowledge management	[136]
	Top management understanding	[81, 136, 137]
	Networking and collaboration competences	[136]

Note: Authors' classification.

Technological capabilities do not consist of any specific technology related knowledge and skills. The closest call to a technological capability is the digital capabilities defined in just a few studies, as shown in Table 2. All of these refer to IT-related capabilities with diverse and hard-to-generalize definitions. For example, one study refers to digital capabilities as the combination of a flexible IT infrastructure and a well-developed information management capability [127]. Another considers digital capabilities as a company's capacity to utilize its available IT resources [31].

Further, one third of studies related to technological capabilities are conceptual. For example, the study by Li and Chan [25] presents an in-depth conceptual model developed for IT departments. This unique study offers companies three sets of capabilities to manage their IT: 1) dynamic digital platforms covering IT infrastructure functionality, flexibility, and integration capability; 2) dynamic IT management consisting of IT deployment, exploration, and exploitation; and 3) dynamic IT knowledge management based on knowledge creation, transfer, and retention.

In another study, Gurbaxani [81] offers a framework for DT consisting of six themes: strategic vision, the culture of innovation, know-how and IP assets, digital capabilities, strategic alignment, and tech assets. The majority of these items are managerial, but digital capabilities and tech assets speak to technical capabilities. For example, digital capability refers to the availability of expertise at both the strategic and technical levels and the level of skill at hand to define and execute digital strategies. Tech assets covers big data, data mining and analysis/data analytics, mobile technologies, cloud computing, and internet and wireless communications. These are deemed sufficient technology assets to implement a strategic vision. However, no details are

supplied regarding the expertise needed to use any specific tech asset. The study simply asks survey respondents to rate their company's position in comparison to rival companies.

The remaining digital capabilities mentioned related to IT, such as a company's big data assets [128], or its ability to undertake information analytics [129] or relational/information processing [130]. Many studies focus on general technologies, but they do not discuss any capabilities associated with specific technologies. For example, Muninger et al. [136] investigated the capabilities needed to use social media as a way to generate innovation, finding three non-technical capabilities companies should build upon: knowledge management, top management understanding, and networking and collaboration.

There are also inconsistencies in the broad categories of capabilities, which further complicates a general understanding of what is required for a successful DT. For example, the authors of one study refer to "managerial capability" [125] when what they really mean is the level of technical knowledge managers have. Further, managerial capability is measured by qualitatively ranking the managers' responses to questions such as "Do you have employees dedicated to the management and/or research of new digital technology for your farm?" (the study concerns agriculture). However, again, the paper provides no details of any specific technologies.

This exclusion of technological capabilities from studies presents an intriguing opportunity for future studies to explore. Even though it is speculative, we think the gap may exist for two main reasons. First, it might be challenging to find common technological capabilities considering the wide range of different technological features for each digital technology. Second, the literature is vague on the definition of technological capabilities. Most articles seem to rely on infrastructure or technological investments as an indication of a company's technological capabilities [25]. But this approach ignores the importance of the soft side of technologies, particularly know-how and intellectual property rights [136].

To fill this knowledge gap in the literature, this study draws on understanding technology management as a set of capabilities [138]. Further, RQ3 asks: What are the AI capabilities enabling DT? Through our third question, we focus on one digital technology, AI, and assess it through the most widely used indicator of technological proficiency: patents. We used network analytics to deconstruct the key technical knowledge associated with a company in this arena. The next section is dedicated to exemplifying how technological capabilities can be derived for different technologies by using patent analysis tools in a case study on AI.

5. Case study: AI patents

By feeding the technological terms extracted from Dataset 3 into our HTT algorithm, we generated the HTT in Figure 4. The weight of edges, in this case, represents the normalized value of the co-occurrences between connected techniques - i.e., how many patents these two techniques appear together in our dataset 3. We set the edge weight threshold to 15 to retain the concise HTT of AI technologies, resulting in 55 nodes. The finalized AI technologies HTT was

used to identify the technological composition, divergence, and convergence of patented AI techniques.

Our hierarchical clustering algorithm partitioned the HTT into seven clusters. As a general summary, this HTT's primary root node is *sensor* with three linked topic clusters: *sensor technologies* #A; *transmission technologies* #B; *AI applications* #C, which owns *robot*, *cloud technology*, and *Internet of Things (IoT)* as three typical AI applications. Derived from cluster #C, *robot* is further broken down into *robotics* (#D) and *robot functions* (#E), *cloud technology* leads a whole cluster with the same name, i.e., *cloud technology* #F. *Neural networks* (#G) is a relatively independent cluster parent node that converges with the node *machine learning*.
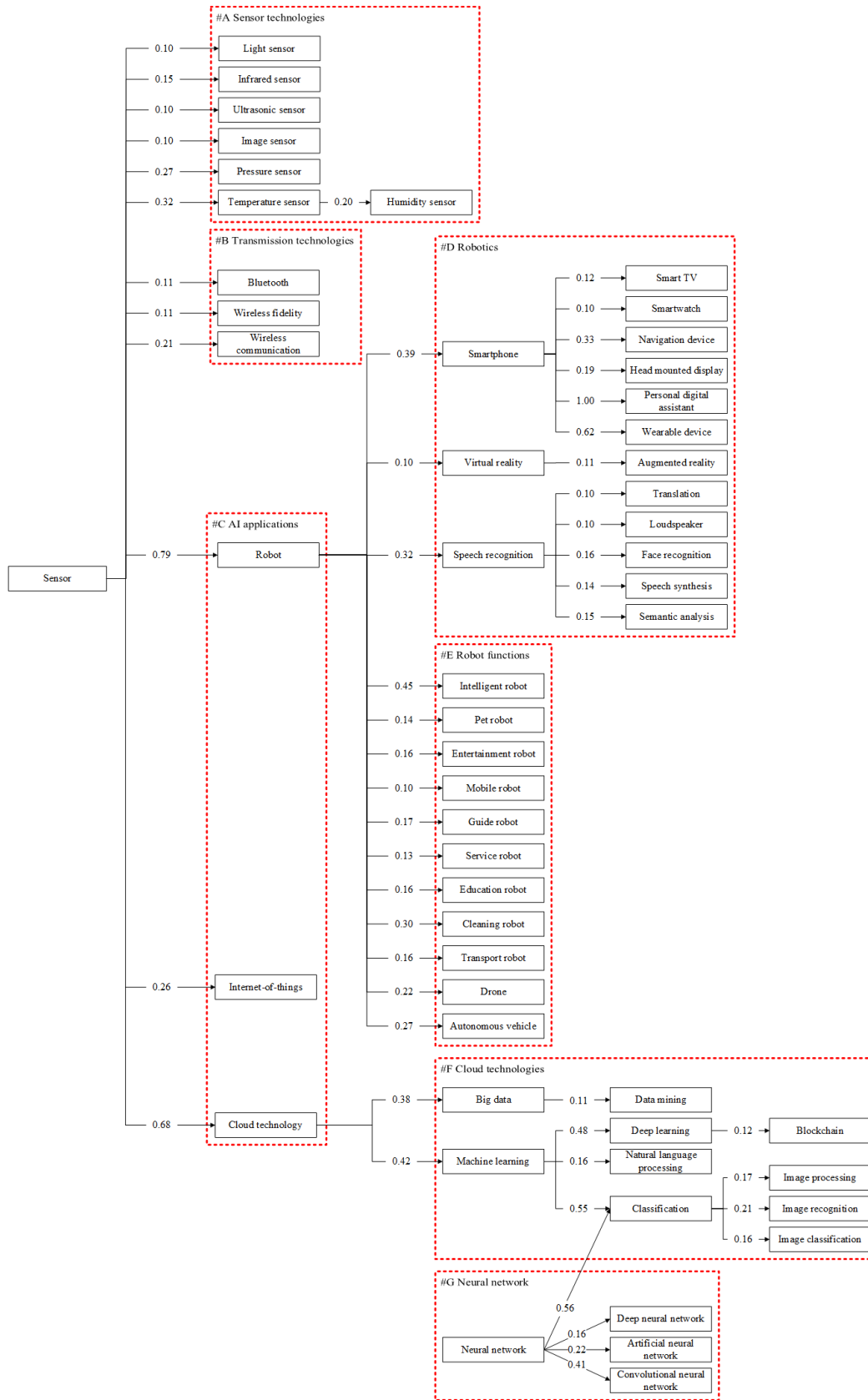
Figure 4. The HTT generated from the AI patents dataset

The status of sensor technology as the root node demonstrates its foundational role in AI. Our patent review also confirmed that sensors were used as a basic data collection module in most of the granted patents. The detailed reviews of each technique cluster are given as follows:

Sensor technologies (#A) consist of various sensors designed to capture different input signals, such as pressure, humidity, images, ultrasonic waves, temperature, infrared light, etc.

Transmission technologies (#B) comprises information and communication technologies used in data transmission modules, such as Bluetooth, wireless fidelity (WIFI), and wireless communication.

AI applications (#C) contains robot, cloud technology and IoT as three representative applications. *Robot* is the largest subordinate node and is further partitioned into *robotics* (#D) and *robot functions* (#E). *Robotics* #D is an interesting cluster housing many innovations that sit at the convergence of multiple other technologies, such as speech recognition, face recognition, smartphone, wearable devices, virtual reality, and augmented reality. *Robot functions* (#E) covers the robots built for various specific uses, including drones, autonomous vehicles, pets, navigation, cleaning, service, education, etc. *IoT* is another application in this cluster but it has no subordinate nodes. Our patent review reveals that current major patents in IoT tend to focus on intelligent hardware control. However, the terms describing pure hardware facilities (like rotating rod, supporting rod, etc.) were not included in the filtered technological terms. Our findings of the emergence of IoT and robotics in AI applications comply with the co-evolving patterns and convergences identified by Katy et al. [139], whose citation analysis indicates that cross-citation between AI, IoT, and robotics have increased dramatically over the last decade.

Cloud technology (#F) is a special AI application that plays a bridge role in connecting AI services with end-users. This cluster involves substantial specific AI algorithms and techniques. By referring to the relevant patents in this cluster, we found that AI algorithms and techniques always involve massive data processing and need cloud technology to provide a computing efficiency solution. From this perspective, cloud technology can be regarded as the prerequisite technology for the product realization of AI. When diving into the subordinate nodes of *cloud technology*, we identify the following technological composition and changes:

- Technological segmentation (#F): *Machine learning* is segmented into *deep learning*, *natural language processing* (NLP), and classification. Classification diverges into image processing, image classification, and image recognition, indicating that most classification tasks are related to image data.
- Technological convergence (#F & #G): machine learning and neural networks together present a technological convergence to classification. This convergence indicates the incorporation of machine learning and neural network in improving the accuracy of classification tasks.

Based on Figure 4, we find AI capabilities can be classified into four levels:

1) Data collection and transmission: the capability to leverage technologies that collect data from the physical world or transfer data within and between product modules. Sensors and ICTs are the representative technologies that realize such capabilities in AI inventions.

2) Bridging: the ability to connect (disparate) end-users with AI products and services/products. Cloud technology is a crucial part of bridging capability since it plays an indispensable role in the deployment and large-scale implementation of advanced AI algorithms.

3) Algorithms: the ability to use AI techniques and algorithms to perform specific business tasks. Typical examples include machine learning, deep learning, big data analysis, neural networks, etc.

4) Applications: the capability to realize mature technological convergences between AI and/or other technologies to provide innovative products. IoT and robots, for example, are two mainstream applications of such a kind.

With this stratification, we can conceptualize the capabilities needed to successfully leverage AI within a DT process as the pyramid shown in Figure 5.
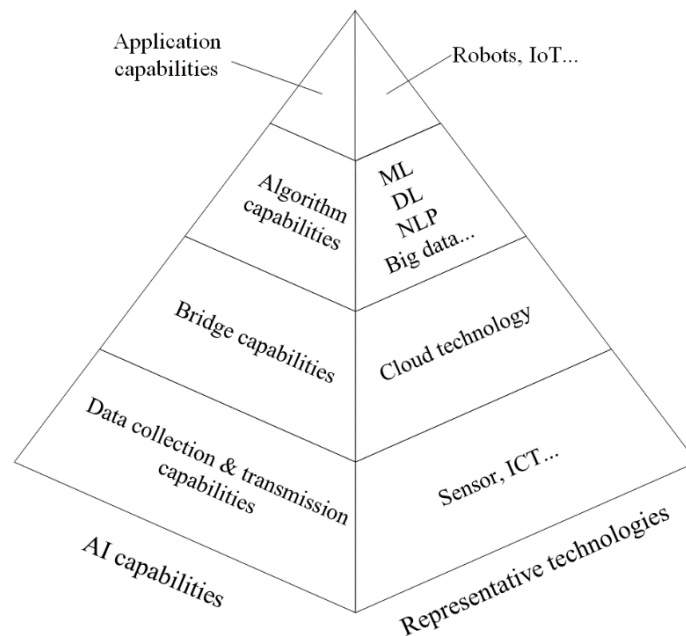
Figure 5. The pyramid of AI capabilities

6. Discussion and conclusions

DT is here to stay, and its revolutionary nature seems to speed up in parallel to rapid changes in digital technologies [13]. Clarifying its definition, observing its evolution, and identifying its enablers could benefit both information systems and technology management disciplines.

Without clear and empirically validated definitions, the DT literature might remain in an adolescent phase of development.

6.1. Key findings

In this study, we exploit methods of intelligent bibliometrics, including scientific evolutionary pathways, HTT analysis, and network analytics, to conduct a set of quantitative analyses as opposed to the qualitative analyses dominating the field. Through these approaches, we address three critical questions in DT:

Q1: What is the definition of DT from a bibliometric perspective?

The SEP analysis advances Vial's definition and solidifies those concepts with identified three stages of dynamic research topic communities. Additionally, the HTT analysis highlights several DT enablers, including digital technologies, digital platforms, and digital capabilities, based on which researchers can further extend DT's definition [1, 24, 25, 28]. This study proposes to define DT as a process that aims to improve an entity by triggering significant changes to its properties through combinations of digital capabilities, technologies, and platforms.

Q2: What are the capabilities enabling DT?

Our literature review, driven by the SEP and HTT analyses, brings a broad range of digital capabilities together into a comprehensive set of categories, as given in Table 2. The studies reviewed highlight the key role of dynamic, technological, and platform capabilities in DT. In addition, our review underscores the negligence of technological capabilities in literature [4, 5, 32]. Hence, we assert that literature fails to develop a comprehensive understanding of DT that is, after all, enabled by digital technologies [128].

Q3: What are the AI capabilities enabling DT?

Having decomposed the hierarchical technologies from AI patents, we propose the pyramid model of four major capabilities illustrated in Figure 5: data collection & transmission, bridging, algorithms, and applications. By focusing on AI that is an exemplar of digital technology, we underline the technology-level capabilities that could make a difference in transforming organizations [81]. This nuanced approach calls researchers to do similar exercises for other digital technologies that could help to build knowledge on actual and varied interactions among digital technologies and DT.

6.2. Implications for theory

This paper has three academic contributions to the DT literature that seems an emerging field [1]. First, it offers a structured investigation of the conceptualization of DT by reviewing prior research through the SEP and HTT analyses. The findings present a succinct quantitative account of the breadth and depth of DT research in the last decade. It presents the evolution of the field and introduces hierarchical clusters of themes within DT. Findings indicate how the research in

the DT literature evolved from the fundamental concepts such as digitization [22] towards the implementation of DT like changes in business models [49] and then to dive-in studies into individual digital technologies such as AI [65, 66]. The hierarchical clusters pointed to four key themes of the extant literature: Digitization, ICTs, digital technologies, and digitization processes. Hence, we believe our paper provides an overview of the literature.

Second, the study expands the understanding of the DT concept [1]. Thanks to the solid quantitative evidence generated from the SEP and HTT, the study highlights DT as a process of digital capabilities, technologies, and platforms.

Third, the bibliometric analysis reveals three main capabilities enabling DT: dynamic, technological, and platform capabilities. Further, the paper introduces an approach to identify technological capabilities through patent analysis. The decomposition of hierarchical technologies from AI patents has resulted in matching capabilities for each sub-technology within AI. This approach offers the literature a tool to conduct a more in-depth analysis of technological capabilities associated with individual digital technologies [11]. Hence, our findings might also be interesting for researchers discussing the dynamic capabilities theory [6, 6, 8, 29, 30].

6.3. Implications for practice

As our findings indicate, the current literature is patchy and incomplete in understanding the capabilities needed for a successful DT [1, 9, 33]. Therefore, our study offers one major contribution to practice by bringing diverse capabilities together under three major categories: dynamic, technological, and platform capabilities. Offering such a structured list of capabilities helps to make managers aware of some of the core resources and competencies that will likely be helpful on their DT journeys. Further, we overcome the simplified understanding of technological capabilities as a set of managerial capabilities by offering capabilities enabling AI. Our in-depth analysis draws managers' attention to consider both general and specific capabilities required for their transformation activities arising from the adoption of AI. This nuanced understanding of capabilities and DT could improve decision-making on how managers could develop their managerial and technological capabilities. As shown in a study [32], the relationship between technology management capabilities and technology capabilities affects the overall performance of companies.

6.4. Limitations and future directions

Nevertheless, there are some limitations to our work. From a methodological perspective, two major limitations stem from the HTT algorithm: 1) Although the algorithm reveals the hierarchies hidden in a co-occurrence network intuitively, the maximum spanning tree algorithm it includes inevitably causes some loss of information, some of which might include potential emerging topic interactions. To minimize this loss, in the future, we anticipate adding a pre-clustering process ahead of the tree generation step. This change would preserve important

secondary edges in the network. 2) Based on expert judgments, we empirically selected two thresholds for the tree generation process to remove trivial branches. In the future, we plan to improve the method's adaptability with functionality that automatically generates a recommended threshold value range based on the results of a sensitivity test.

From a theoretical perspective, there are two major limitations. First, this study is a conceptual paper drawing from the extant literature and sheds light on the definition of DT, its evolution, and the capabilities that enable it. Future studies might gather empirical evidence regarding the relationship between capabilities and DT. This research avenue might investigate the relationship between capabilities enabling a general DT or a specific digital technology. For example, researchers might conduct an in-depth analysis of relationships among different capabilities and their impact on the outcome of DT specifically for AI. Such a line of research could bring abundant knowledge on the area of AI-enabled capabilities as a whole picture.

Second, our analysis does not analyze the role of diverse contexts, such as different industries, countries, and organization types. Again, this limitation makes us invite researchers for another line of research about how capabilities enabling DT might differ based on contextual factors in practice. For example, a recent study in this journal [140] has developed a transdisciplinary system design to describe a generic model applicable to any industry sector going through a DT. The study is applied to two industry sectors: Education and Retail, highlighting how the generic model changes in each context. Hence, we think that future studies could generate coherent efforts to garner novel and relevant knowledge in DT by delving into diverse contexts, including industries and technologies.

Acknowledgments

References

[1]   G. Vial, "Understanding digital transformation: A review and a research agenda," *The Journal of Strategic Information Systems,* vol. 28, no. 2, pp. 118-144, 2019.

[2]   J. Reis, M. Amorim, N. Melão, and P. Matos, "Digital transformation: a literature review and guidelines for future research," in *World Conference on Information Systems and Technologies*, 2018: Springer, pp. 411-421.

[3]   Y. Zhang, Y. Guo, X. Wang, D. Zhu, and A. L. Porter, "A hybrid visualisation model for technology roadmapping: Bibliometrics, qualitative methodology and empirical study," *Technology Analysis & Strategic Management,* vol. 25, no. 6, pp. 707-724, 2013.

[4]   A. S. Bharadwaj, "A resource-based perspective on information technology capability and firm performance: An empirical investigation," *MIS Quarterly,* pp. 169-196, 2000.

[5]   H.-C. Chae, C. E. Koh, and V. R. Prybutok, "Information technology capability and firm performance: Contradictory findings and their possible causes," *MIS Quarterly,* vol. 38, no. 1, pp. 305-326, 2014.

[6]   K. M. Eisenhardt and J. A. Martin, "Dynamic capabilities: What are they?," *Strategic Management Journal,* vol. 21, no. 10-11, pp. 1105-1121, 2000.

[7]   C. E. Helfat *et al.*, *Dynamic capabilities: Understanding strategic change in organizations*. John Wiley & Sons, 2009.

[8]   D. J. Teece, "Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance," *Strategic Management Journal,* vol. 28, no. 13, pp. 1319-1350, 2007.

[9]   K. S. Warner and M. Wäger, "Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal," *Long Range Planning,* vol. 52, no. 3, pp. 326-349, 2019.

[10]   D. Rotolo, D. Hicks, and B. R. Martin, "What is an emerging technology?," *Research policy,* vol. 44, no. 10, pp. 1827-1843, 2015.

[11]   J. Kietzmann and L. F. Pitt, "Artificial intelligence and machine learning: What managers need to know," *Business Horizons,* vol. 63, no. 2, pp. 131-133, 2020.

[12]   P. K. Muhuri, A. K. Shukla, and A. Abraham, "Industry 4.0: A bibliometric analysis and detailed overview," *Engineering applications of artificial intelligence,* vol. 78, pp. 218-235, 2019.

[13]   WEF, *Digital Transformation Initiative Maximizing the Return on Digital Investments*. World Economic Forum and Accenture, WEF, Ed., 2018. http://www3.weforum.org/docs/DTI_Maximizing_Return_Digital_WP.pdf

[14]   S. Nambisan, K. Lyytinen, A. Majchrzak, and M. Song, "Digital Innovation Management: Reinventing innovation management research in a digital world," *MIS Quarterly,* vol. 41, no. 1, 2017.

[15]   Y. Zhang, G. Zhang, H. Chen, A. L. Porter, D. Zhu, and J. Lu, "Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research," *Technological Forecasting and Social Change,* vol. 105, pp. 179-191, 2016.

[16]   S. F. Carley, N. C. Newman, A. L. Porter, and J. G. Garner, "An indicator of technical emergence," *Scientometrics,* vol. 115, no. 1, pp. 35-49, 2018.

[17]   M. Wu, Y. Zhang, G. Zhang, and J. Lu, "Exploring the genetic basis of diseases through a heterogeneous bibliometric network: A methodology and case study," *Technological Forecasting and Social Change,* vol. 164, p. 120513, 2020.

[18]   C. V. Fry, X. Cai, Y. Zhang, and C. S. Wagner, "Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research," (in eng), *PLoS One,* vol. 15, no. 7, p. e0236307, 2020, doi: 10.1371/journal.pone.0236307.

[19]   Y. Zhang, Y. Huang, A. L. Porter, G. Zhang, and J. Lu, "Discovering and forecasting interactions in big data research: A learning-enhanced bibliometric study," *Technological Forecasting and Social Change,* vol. 146, pp. 795-807, 2019.

[20]   Y. Zhang, H. Chen, J. Lu, and G. Zhang, "Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016," *Knowledge-Based Systems,* vol. 133, pp. 255-268, 2017.

[21]   Y. Zhang, G. Zhang, D. Zhu, and J. Lu, "Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics," *Journal of the Association for Information Science and Technology,* vol. 68, no. 8, pp. 1925-1939, 2017.

[22]   Y. Yoo, O. Henfridsson, and K. Lyytinen, "Research commentary—the new organizing logic of digital innovation: An agenda for information systems research," *Information Systems Research,* vol. 21, no. 4, pp. 724-735, 2010.

[23]   S. Nambisan, M. Wright, and M. Feldman, "The Digital transformation of innovation and entrepreneurship: Progress, challenges and key themes," *Research Policy,* vol. 48, no. 8, p. 103773, 2019.

[24]   E. Autio, S. Nambisan, L. D. Thomas, and M. Wright, "Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems," *Strategic Entrepreneurship Journal,* vol. 12, no. 1, pp. 72-95, 2018.

[25]   T. C. Li and Y. E. Chan, "Dynamic information technology capability: Concept definition and framework development," *The Journal of Strategic Information Systems,* vol. 28, no. 4, p. 101575, 2019.

[26]   B. Faro, B. Abedin, and D. Cetindamar, "Continuous Transformation of Public Sector Organisations in the Digital Era," 2019.

[27]   D. J. Teece and G. Linden, "Business models, value capture, and the digital enterprise," *Journal of Organization Design,* vol. 6, no. 1, pp. 1-14, 2017.

[28]   M. Subramaniam, B. Iyer, and V. Venkatraman, "Competing in digital ecosystems," *Business Horizons,* vol. 62, no. 1, pp. 83-94, 2019.

[29]   D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strategic Management Journal,* vol. 18, no. 7, pp. 509-533, 1997.

[30]   C. E. Helfat and S. G. Winter, "Untangling dynamic and operational capabilities: Strategy for the (N) ever-changing world," *Strategic Management Journal,* vol. 32, no. 11, pp. 1243-1250, 2011.

[31]   C. Fernandes, J. J. Ferreira, M. L. Raposo, C. Estevão, M. Peris-Ortiz, and C. Rueda-Armengot, "The dynamic capabilities perspective of strategic management: A co-citation analysis," *Scientometrics,* vol. 112, no. 1, pp. 529-555, 2017.

[32]   W. Wu, Z. Liang, Q. Zhang, and H. Zhang, "Coupling relationships and synergistic mechanisms between technology management capability and technological capability in product innovation: A simulation study," *Technology Analysis & Strategic Management,* pp. 1-15, 2020.

[33]   X. Liu, R. Torres de Oliveira, M. Indulska, and M.-L. Verreynne, "Towards a Digital Capability Framework," in *Academy of Management Proceedings*, 2020, vol. 2020, no. 1: Academy of Management Briarcliff Manor, NY 10510, p. 18505.

[34]   J. Sears and G. Hoetker, "Technological overlap, technological capabilities, and resource recombination in technological acquisitions," *Strategic Management Journal,* vol. 35, no. 1, pp. 48-67, 2014.

[35]   W. Wang, A. Mahmood, C. Sismeiro, and N. Vulkan, "The evolution of equity crowdfunding: Insights from co-investments of angels and the crowd," *Research Policy,* vol. 48, no. 8, p. 103727, 2019.

[36]   D. J. Price, *Little science, big science... and beyond*. Columbia University Press New York, 1986.

[37]   A. Porter, Y. Zhang, Y. Huang, and M. Wu, "Tracking and mining the COVID-19 research literature," *Frontiers in Research Metrics and Analytics,* 2020, doi: 10.3389/frma.2020.594060.

[38]   Y. Zhang, A. L. Porter, S. Cunningham, D. Chiavetta, and N. Newman, "Parallel or Intersecting Lines? Intelligent Bibliometrics for Investigating the Involvement of Data Science in Policy Analysis," *IEEE Transactions on Engineering Management,* pp. 1-13, 2020, doi: 10.1109/TEM.2020.2974761.

[39]   Y. Zhang *et al.*, "Does deep learning help topic extraction? A kernel k-means clustering method with word embedding," *Journal of Informetrics,* vol. 12, no. 4, pp. 1099-1117, 2018.

[40]   J. Hou, X. Yang, and C. Chen, "Emerging trends and new developments in information science: A document co-citation analysis (2009–2016)," *Scientometrics,* vol. 115, no. 2, pp. 869-892, 2018.

[41]   C. Wartena, R. Brussee, and W. Slakhorst, "Keyword extraction using word co-occurrence," in *2010 Workshops on Database and Expert Systems Applications*, 2010: IEEE, pp. 54-58.

[42]   G. Colavizza and M. Franceschet, "Clustering citation histories in the Physical Review," *Journal of Informetrics,* vol. 10, no. 4, pp. 1037-1051, 2016.

[43]   S. Ravikumar, A. Agrahari, and S. N. Singh, "Mapping the intellectual structure of scientometrics: A co-word analysis of the journal Scientometrics (2005–2010)," *Scientometrics,* vol. 102, no. 1, pp. 929-955, 2015.

[44]   J. Guo, X. Wang, Q. Li, and D. Zhu, "Subject–action–object-based morphology analysis for determining the direction of technological change," *Technological Forecasting and Social Change,* vol. 105, pp. 27-40, 2016.

[45]   Y. Zhang, X. Wang, G. Zhang, and J. Lu, "Predicting the dynamics of scientific activities: A diffusion-based network analytic methodology," *Proceedings of the Association for Information Science and Technology,* vol. 55, no. 1, pp. 598-607, 2018, doi: 10.1002/pra2.2018.14505501065.

[46]   R. N. Kostoff, R. Boylan, and G. R. Simons, "Disruptive technology roadmaps," *Technological Forecasting and Social Change,* vol. 71, no. 1-2, pp. 141-159, 2004.

[47]   C. Wang *et al.*, "A phrase mining framework for recursive construction of a topical hierarchy," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013, pp. 437-445.

[48]   J. Shang, X. Zhang, L. Liu, S. Li, and J. Han, "Nettaxo: Automated topic taxonomy construction from text-rich network," in *Proceedings of the Web Conference 2020*, 2020, pp. 1908-1919.

[49]   D. Schallmo, C. A. Williams, and L. Boardman, "Digital transformation of business models—best practice, enablers, and roadmap," *International Journal of Innovation Management,* vol. 21, no. 08, p. 1740014, 2017.

[50]   P. Jacsó, "Metadata mega mess in Google Scholar," *Online Information Review,* 2010.

[51]   A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. D. López-Cózar, "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations," *Scientometrics,* vol. 126, no. 1, pp. 871-906, 2021.

[52]   A.-W. Harzing and S. Alakangas, "Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison," *Scientometrics,* vol. 106, no. 2, pp. 787-804, 2016.

[53]   A. Aghaei Chadegani *et al.*, "A comparison between two main academic literature collections: Web of Science and Scopus databases," *Asian Social Science,* vol. 9, no. 5, pp. 18-26, 2013.

[54]   C. López-Illescas, F. de Moya-Anegón, and H. F. Moed, "Coverage and citation impact of oncological journals in the Web of Science and Scopus," *Journal of Informetrics,* vol. 2, no. 4, pp. 304-316, 2008.

[55]   T. Mahlia *et al.*, "Patent landscape review on biodiesel production: Technology updates," *Renewable and Sustainable Energy Reviews,* vol. 118, p. 109526, 2020.

[56]   Y. Zhang, A. L. Porter, Z. Hu, Y. Guo, and N. C. Newman, ""Term clumping" for technical intelligence: A case study on dye-sensitized solar cells," *Technological Forecasting and Social Change,* vol. 85, pp. 26-39, 2014.

[57]   L. Fleming, "Recombinant uncertainty in technological search," *Management Science,* vol. 47, no. 1, pp. 117-132, 2001.

[58]   L. Fleming and O. Sorenson, "Science as a map in technological search," *Strategic Management Journal,* vol. 25, no. 8-9, pp. 909-928, 2004.

[59]   M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences,* vol. 103, no. 23, pp. 8577-8582, 2006.

[60]   S. Pemmaraju and S. Skiena, *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica®.* Cambridge University Press, 2003.

[61]   A. Hein *et al.*, "Digital platform ecosystems," *Electronic Markets,* pp. 1-12, 2019.

[62]   J. S. Butler, R. Garg, and B. Stephens, "Social networks, funding, and regional advantages in technology entrepreneurship: an empirical analysis," *Information Systems Research,* vol. 31, no. 1, pp. 198-216, 2020.

[63]   D. Zhukov, T. Khvatova, S. Lesko, and A. Zaltcman, "Managing social networks: Applying the percolation theory methodology to understand individuals' attitudes and moods," *Technological Forecasting and Social Change,* vol. 129, pp. 297-307, 2018.

[64]   P. Neirotti and D. Pesce, "ICT-based innovation and its competitive outcome: The role of information intensity," *European Journal of Innovation Management,* 2019.

[65]   X. Yang, D. Cao, J. Chen, Z. Xiao, and A. Daowd, "AI and IoT-based collaborative business ecosystem: A case in Chinese fish farming industry," *International Journal of Technology Management,* vol. 82, no. 2, pp. 151-171, 2020.

[66]   J. K.-U. Brock and F. Von Wangenheim, "Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence," *California Management Review,* vol. 61, no. 4, pp. 110-134, 2019.

[67]   C. Tumelero, R. Sbragia, and S. Evans, "Cooperation in R & D and eco-innovations: The role in companies' socioeconomic performance," *Journal of Cleaner Production,* vol. 207, pp. 1138-1149, 2019.

[68]   R. Nicolescu, M. Huth, P. Radanliev, and D. De Roure, "Mapping the values of IoT," *Journal of Information Technology,* vol. 33, no. 4, pp. 345-360, 2018.

[69]   S. Bayer, H. Gimpel, and D. Rau, "IoT-commerce-opportunities for customers through an affordance lens," *Electronic Markets,* pp. 1-24, 2020.

[70]   A. Young, L. Selander, and E. Vaast, "Digital organizing for social impact: Current insights and future research avenues on collective action, social movements, and digital technologies," *Information and Organization,* vol. 29, no. 3, p. 100257, 2019.

[71]   E. Danneels, "Disruptive technology reconsidered: A critique and research agenda," *Journal of Product Innovation Management,* vol. 21, no. 4, pp. 246-258, 2004.

[72]   M. Ardolino, M. Rapaccini, N. Saccani, P. Gaiardelli, G. Crespi, and C. Ruggeri, "The role of digital technologies for the service transformation of industrial companies," *International Journal of Production Research,* vol. 56, no. 6, pp. 2116-2132, 2018.

[73]   J. L. Hartley and W. J. Sawaya, "Tortoise, not the hare: Digital transformation of supply chain business processes," *Business Horizons,* vol. 62, no. 6, pp. 707-715, 2019.

[74]   Z. Wang, H. Zhao, and Y. Wang, "Social networks in marketing research 2001–2014: A co-word analysis," *Scientometrics,* vol. 105, no. 1, pp. 65-82, 2015.

[75]   B. K. Chae, "A General framework for studying the evolution of the digital innovation ecosystem: The case of big data," *International Journal of Information Management,* vol. 45, pp. 83-94, 2019.

[76]   M. Afshar Ali, K. Alam, and B. Taylor, "Do social exclusion and remoteness explain the digital divide in Australia? Evidence from a panel data estimation approach," *Economics of Innovation and New Technology,* vol. 29, no. 6, pp. 643-659, 2020.

[77]   S. C. Srivastava and G. Shainesh, "Bridging the service divide through digitally enabled service innovations: Evidence from indian healthcare service providers," *MIS Quarterly,* vol. 39, no. 1, pp. 245-267, 2015.

[78]   O. Henfridsson, L. Mathiassen, and F. Svahn, "Managing technological change in the digital age: The role of architectural frames," *Journal of Information Technology,* vol. 29, no. 1, pp. 27-43, 2014.

[79]   S. Øiestad and M. M. Bugge, "Digitisation of publishing: Exploration based on existing business models," *Technological Forecasting and Social Change,* vol. 83, pp. 54-65, 2014.

[80]   D. Shin, "How do users experience the interaction with an immersive screen?," *Computers in Human Behavior,* vol. 98, pp. 302-310, 2019.

[81]   V. Gurbaxani and D. Dunkle, "Gearing Up For Successful Digital Transformation," *MIS Quarterly Executive,* vol. 18, no. 3, 2019.

[82]   A. Pagoropoulos, A. Maier, and T. C. McAloone, "Assessing transformational change from institutionalising digital capabilities on implementation and development of Product-Service Systems: Learnings from the maritime industry," *Journal of Cleaner Production,* vol. 166, pp. 369-380, 2017.

[83]   K. Demeter, D. Losonci, and J. Nagy, "Road to digital manufacturing–a longitudinal case-based analysis," *Journal of Manufacturing Technology Management,* 2020.

[84] A. Jantunen, A. Tarkiainen, S. Chari, and P. Oghazi, "Dynamic capabilities, operational changes, and performance outcomes in the media industry," *Journal of Business Research,* vol. 89, pp. 251-257, 2018.

[85] J. Karimi and Z. Walter, "The role of dynamic capabilities in responding to digital disruption: A factor-based study of the newspaper industry," *Journal of Management Information Systems,* vol. 32, no. 1, pp. 39-81, 2015.

[86] K. North, N. Aramburu, and O. J. Lorenzo, "Promoting digitally enabled growth in SMEs: A framework proposal," *Journal of Enterprise Information Management,* 2019.

[87] Y. L. Antonucci, A. Fortune, and M. Kirchmer, "An examination of associations between business process management capabilities and the benefits of digitalization: All capabilities are not equal," *Business Process Management Journal,* 2020.

[88] D. Nylén and J. Holmström, "Digital innovation strategy: A framework for diagnosing and improving digital product and service innovation," *Business Horizons,* vol. 58, no. 1, pp. 57-67, 2015.

[89] R. Pershina, B. Soppe, and T. M. Thune, "Bridging analog and digital expertise: Cross-domain collaboration and boundary-spanning tools in the creation of digital innovation," *Research Policy,* vol. 48, no. 9, p. 103819, 2019.

[90] D. Trabucchi and T. Buganza, "Data-driven innovation: Switching the perspective on Big Data," *European Journal of Innovation Management,* 2019.

[91] M. Candi and A. Beltagui, "Effective use of 3D printing in the innovation process," *Technovation,* vol. 80, pp. 63-73, 2019.

[92] K. Trantopoulos, G. von Krogh, M. W. Wallin, and M. Woerter, "External knowledge and information technology: Implications for process innovation performance," *MIS Quarterly,* vol. 41, no. 1, pp. 287-300, 2017.

[93] F. Li, "The digital transformation of business models in the creative industries: A holistic framework and emerging trends," *Technovation,* vol. 92, p. 102012, 2020.

[94] C. Loebbecke and A. Picot, "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda," *The Journal of Strategic Information Systems,* vol. 24, no. 3, pp. 149-157, 2015.

[95] A. Bharadwaj, O. A. El Sawy, P. A. Pavlou, and N. Venkatraman, "Digital business strategy: Toward a next generation of insights," *MIS Quarterly,* pp. 471-482, 2013.

[96] A. Correani, A. De Massis, F. Frattini, A. M. Petruzzelli, and A. Natalicchio, "Implementing a digital strategy: Learning from the experience of three digital transformation projects," *California Management Review,* vol. 62, no. 4, pp. 37-56, 2020.

[97] J. Frishammar, A. Richtnér, A. Brattström, M. Magnusson, and J. Björk, "Opportunities and challenges in the new innovation landscape: Implications for innovation auditing and innovation management," *European Management Journal,* vol. 37, no. 2, pp. 151-164, 2019.

[98] M. Kohtamäki, V. Parida, P. C. Patel, and H. Gebauer, "The relationship between digitalization and servitization: The role of servitization in capturing the financial potential of digitalization," *Technological Forecasting and Social Change,* vol. 151, p. 119804, 2020.

[99] J. S. Black and P. van Esch, "AI-enabled recruiting: What is it and how should a manager use it?," *Business Horizons,* vol. 63, no. 2, pp. 215-226, 2020.

[100] J. J. Ferreira, C. I. Fernandes, and F. A. Ferreira, "To be or not to be digital, that is the question: Firm innovation and performance," *Journal of Business Research,* vol. 101, pp. 583-590, 2019.

[101] J. Björkdahl, "Strategies for digitalization in manufacturing firms," *California Management Review,* p. 0008125620920349, 2020.

[102] E. Pessot, A. Zangiacomi, C. Battistella, V. Rocchi, A. Sala, and M. Sacco, "What matters in implementing the factory of the future," *Journal of Manufacturing Technology Management,* 2020.

[103] C. Alaimo and J. Kallinikos, "Computing the everyday: Social media as data platforms," *The Information Society,* vol. 33, no. 4, pp. 175-191, 2017.

[104] C. E. Helfat and R. S. Raubitschek, "Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems," *Research Policy,* vol. 47, no. 8, pp. 1391-1399, 2018.

[105] C. L. Garay-Rondero, J. L. Martinez-Flores, N. R. Smith, S. O. C. Morales, and A. Aldrette-Malacara, "Digital supply chain model in Industry 4.0," *Journal of Manufacturing Technology Management,* 2019.

[106] A. Ghadge, M. E. Kara, H. Moradlou, and M. Goswami, "The impact of Industry 4.0 implementation on supply chains," *Journal of Manufacturing Technology Management,* 2020.

[107] D. Ivanov, A. Dolgui, and B. Sokolov, "The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics," *International Journal of Production Research,* vol. 57, no. 3, pp. 829-846, 2019.

[108] G. Culot, G. Orzes, M. Sartor, and G. Nassimbeni, "The future of manufacturing: A Delphi-based scenario analysis on Industry 4.0," *Technological Forecasting and Social Change,* vol. 157, p. 120092, 2020.

[109] M. Ghobakhloo and M. Fathi, "Corporate survival in Industry 4.0 era: The enabling role of lean-digitized manufacturing," *Journal of Manufacturing Technology Management,* 2019.

[110] D. Horváth and R. Z. Szabó, "Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?," *Technological Forecasting and Social Change,* vol. 146, pp. 119-132, 2019.

[111] T. Zheng, M. Ardolino, A. Bacchetti, M. Perona, and M. Zanardini, "The impacts of Industry 4.0: A descriptive survey in the Italian manufacturing sector," *Journal of Manufacturing Technology Management,* 2019.

[112] R. Agarwal, G. Gao, C. DesRoches, and A. K. Jha, "Research commentary—The digital transformation of healthcare: Current status and the road ahead," *Information Systems Research,* vol. 21, no. 4, pp. 796-809, 2010.

[113] I. Holeman and D. Kane, "Human-centered design for global health equity," *Information Technology for Development,* vol. 26, no. 3, pp. 477-505, 2020.

[114] A. Ghazawneh and O. Henfridsson, "A paradigmatic analysis of digital application marketplaces," *Journal of Information Technology,* vol. 30, no. 3, pp. 198-208, 2015.

[115] R. D. Wang and C. D. Miller, "Complementors' engagement in an ecosystem: A study of publishers' e-book offerings on Amazon Kindle," *Strategic Management Journal,* vol. 41, no. 1, pp. 3-26, 2020.

[116] S. Yablonsky, "A multidimensional platform ecosystem framework," *Kybernetes,* 2020.

[117] F. Bienhaus and A. Haddud, "Procurement 4.0: Factors influencing the digitisation of procurement and supply chains," *Business Process Management Journal,* 2018.

[118] A. G. Frank, L. S. Dalenogare, and N. F. Ayala, "Industry 4.0 technologies: Implementation patterns in manufacturing companies," *International Journal of Production Economics,* vol. 210, pp. 15-26, 2019.

[119] I. Castelo-Branco, F. Cruz-Jesus, and T. Oliveira, "Assessing Industry 4.0 readiness in manufacturing: Evidence for the European Union," *Computers in Industry,* vol. 107, pp. 22-32, 2019.

[120] A. Hidalgo, S. Gabaly, G. Morales-Alonso, and A. Urueña, "The digital divide in light of sustainable development: An approach through advanced machine learning techniques," *Technological Forecasting and Social Change,* vol. 150, p. 119754, 2020.

[121] E. Stoeckli, C. Dremel, and F. Uebernickel, "Exploring characteristics and transformational capabilities of InsurTech innovations to understand insurance value creation in a digital world," *Electronic Markets,* vol. 28, no. 3, pp. 287-305, 2018.

[122] G. S. Day and P. J. Schoemaker, "Adapting to fast-changing markets and technologies," *California Management Review,* vol. 58, no. 4, pp. 59-77, 2016.

[123] H.-F. Lin, J.-Q. Su, and A. Higgins, "How dynamic capabilities affect adoption of management innovations," *Journal of Business Research,* vol. 69, no. 2, pp. 862-876, 2016.

[124] Q. Sun, C. Wang, Y. Zhou, L. Zuo, and J. Tang, "Dominant platform capability, symbiotic strategy and the construction of "Internet+ WEEE collection" business ecosystem: A comparative study of two typical cases in China," *Journal of Cleaner Production,* vol. 254, p. 120074, 2020.

[125] M. C. Annosi, F. Brunetta, A. Monti, and F. Nati, "Is the trend your friend? An analysis of technology 4.0 investment decisions in agricultural SMEs," *Computers in Industry,* vol. 109, pp. 59-71, 2019.

[126] L. Li, F. Su, W. Zhang, and J. Y. Mao, "Digital transformation by SME entrepreneurs: A capability perspective," *Information Systems Journal,* vol. 28, no. 6, pp. 1129-1157, 2018.

[127] N. Levallet and Y. E. Chan, "Role of digital capabilities in unleashing the power of managerial improvisation," *MIS Quarterly Executive,* vol. 17, no. 1, 2018.

[128] C. Dremel, J. Wulf, M. M. Herterich, J.-C. Waizmann, and W. Brenner, "How AUDI AG established big data analytics in its digital transformation," *MIS Quarterly Executive,* vol. 16, no. 2, 2017.

[129] Y. Park and S. Mithas, "Organized complexity of digital business strategy: a configurational perspective," *MIS Quarterly,* vol. 44, no. 1, 2020.

[130] T. Saldanha, S. Mithas, and M. S. Krishnan, "Leveraging customer involvement for fueling innovation: The role of relational and analytical information processing capabilities," *MIS Quarterly,* vol. 41, no. 1, pp. 267-286, 2017.

[131] J. Ukko, M. Nasiri, M. Saunila, and T. Rantala, "Sustainability strategy as a moderator in the relationship between digital business strategy and financial performance," *Journal of Cleaner Production,* vol. 236, p. 117626, 2019.

[132] S. Lobo and J. Whyte, "Aligning and Reconciling: Building project capabilities for digital delivery," *Research Policy,* vol. 46, no. 1, pp. 93-107, 2017.

[133] G. L. Tortorella, A. M. C. Vergara, J. A. Garza-Reyes, and R. Sawhney, "Organizational learning paths based upon industry 4.0 adoption: An empirical study with Brazilian

manufacturers," *International Journal of Production Economics,* vol. 219, pp. 284-294, 2020.

[134] P. Setia, P. Setia, V. Venkatesh, and S. Joglekar, "Leveraging digital technologies: How information quality leads to localized capabilities and customer service performance," *MIS Quarterly,* pp. 565-590, 2013.

[135] A. Szalavetz, "Industry 4.0 and capability development in manufacturing subsidiaries," *Technological Forecasting and Social Change,* vol. 145, pp. 384-395, 2019.

[136] M.-I. Muninger, W. Hammedi, and D. Mahr, "The value of social media for innovation: A capability perspective," *Journal of Business Research,* vol. 95, pp. 116-127, 2019.

[137] O. A. El Sawy, P. Kræmmergaard, H. Amsinck, and A. L. Vinther, "How LEGO built the foundations and enterprise capabilities for digital leadership," *MIS Quarterly Executive,* vol. 15, no. 2, 2016.

[138] D. Cetindamar, R. Phaal, and D. Probert, *Technology management: Activities and tools.* Macmillan International Higher Education, 2016.

[139] K. Börner *et al.*, "Mapping the co-evolution of artificial intelligence, robotics, and the internet of things over 20 years (1998-2017)," (in eng), *PLoS One,* vol. 15, no. 12, p. e0242984, 2020, doi: 10.1371/journal.pone.0242984.

[140] M. A. Hashmi, J. P. T. Mo, J. P. T., and R. C. Beckett, " Transdisciplinary systems approach to realization of digital transformation,"*Advanced Engineering Informatics*, vol. 49, 101316, 2021, doi:10.1016/j.aei.2021.101316.

Appendix A. High-quality journal list

Journal Names

Technological forecasting and Social Change	*Journal of The Association for Information Science and Technology*	*Journal of Management Information Systems*
Computers in Human Behavior	*California Management Review*	*Kybernetes*
Journal of Business Research	*Information Society*	*MIT Sloan Management Review*
Electronic Markets	*Interacting with Computers*	*Organization Studies*
Mis Quarterly	*Journal of The Association for Information Systems*	*Strategic Management Journal*
Government Information Quarterly	*Management Decision*	*Urban Studies*
Journal of Cleaner Production	*Organization Science*	*Economics of Innovation and New Technology*
Computers in Industry	*Creativity and Innovation Management*	*European Planning Studies*
International Journal of Information Management	*Futures*	*Industry and Innovation*
Journal of Manufacturing Technology Management	*Information Economics and Policy*	*Production and Operations Management*
Mis Quarterly Executive	*International Journal of Production Economics*	*Decision Support Systems*
Information Systems Research	*Journal of Product Innovation Management*	*Economy and Society*
Telematics and informatics	*Management Science*	*European Management Journal*
Business Horizons	*Scientometrics*	*Information Sciences*
Business Process Management Journal	*Ethics and Information Technology*	*Journal of International Business Studies*
European Journal of Information Systems	*European Journal of Cultural Studies*	*Journal of Management Studies*
Journal of Information Technology	*European Journal of Innovation Management*	*Organization*
Research Policy	*Expert Systems with Applications*	*Small Business Economics*
Technology in Society	*Information Development*	*Tec novation*
International Journal of Production Research	*Information Systems Frontiers*	*IEEE Transactions on Industry Applications*
Information and Organization	*Information Technology for Development*	*International Entrepreneurship and Management Journal*
Information Systems Journal	*International Journal of Technology*	*Journal of Knowledge Management*

Management

Journal of Strategic Information Systems	*Journal of Enterprise Information Management*

Appendix B. The list of papers related to capabilities associated with DT

Source	Type of capabilities	Tech.	Methodology
Antonucci et al. 2020	Business process management capabilities	IT	Survey
Demeter et al, 2019	Sensing, absorptive capacity, integrative, relational	IoT	Case
Sun et al., 2020	Dominant platform capability: digital and network capability	IT	Case
Park, Y. & Mithas, 2020	Digital and non-digital capabilities include information analytics, leadership, strategic planning, customer focus, human resource focus, and process management	IT	Survey
Pessot et al., 2020	Managerial capabilities	Industry 4	Survey & interviews
North et al., 2020	Sensing, seizing and transforming	IT	Survey
Tortorella et al., 2020	Organizational Learning capabilities	Industry 4	Survey
Ukko et al., 2019	Managerial capability and operational capability	Digital tech.	Survey
Gurbaxani, 2019	Strategic vision, culture of innovation, know-how & IP assets, digital capabilities, strategic alignment, tech assets	NA	Survey
Szalavetz, 2019	R&D capabilities, production capability	Industry 4	Interview
Ferreira et al., 2019	Innovative capability	NA	Survey
Annosi et al., 2019	Organizational capabilities (Managerial Capabilities)	Industry 4	Survey
Muninger et al., 2019	Knowledge management, Top management understanding, Networking & collaboration competences	Social media	Theoretical
Li et al., 2018	Dynamic managerial and organizational capabilities	Digital platform	Case
Helfat & Raubitschek, 2018	Innovation capabilities, environmental scanning and sensing capabilities, and integrative capabilities for ecosystem orchestration	Digital platform	Theoretical
Jantunen et al., 2018	Dynamic capabilities	NA	Case
Stoeckli, 2018	Transformational capabilities	IT	Case
Levallet & Chan, 2018	Digital capabilities	NA	Case
Ardolino et al., 2018	Digital capabilities	IoT, cloud, analytics	Case
Pagoropoulos et al., 2017	Digitization and the institutionalization of digital capabilities	NA	Action research
Fernandes et al., 2017	Digital Capabilities, Knowledge Capabilities, Absorptive Capabilities, Strategic Capabilities and Resources	NA	Bibliometric study
Dremel et al., 2017	Big data analytics capabilities	Big data	Case
Saldanha et al., 2017	Relational and analytical information processing capability	IT	Archival data
Lobo & Whyte, 2017	Project capabilities: (1) align the project set-up with the firm's existing capabilities and (2) reconcile differing agendas and capabilities in collaborating firms across the project ecology	IT	Case
Day&Schoemaker, 2016	Sensing, seizing and transforming	NA	Case, theoretical
El Sawy et al., 2016	Capabilities for digital leadership	IT	Case
Karimi & Walter, 2015	Digital platform capabilities	IT	Case
Pankaj et al., 2013	Customer service capabilities	NA	Survey
Li & Chan, 2019	Dynamic IT capabilities	IT	Theoretical
Warner & Wäger,2019	Digital dynamic capabilities	NA	Case
Lin et al., 2016	Sensing capability, absorptive capacity, integrative capability, relational capability	NA	Survey