

A firm's creation of proprietary knowledge linked to the knowledge spilled over from its research publications: the case of artificial intelligence

Su Jung Jee ^{1,2,3,*} and So Young Sohn ²

¹University of Bradford, Faculty of Management, Law and Social Sciences, Richmond Road, Bradford BD7 1DP, UK. e-mail: s.j.jee@bradford.ac.uk, ²Department of Industrial Engineering, Yonsei University, 134 Shinchon-dong, Seoul 120-749, Republic of Korea. e-mail: sohns@yonsei.ac.kr and ³Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK.

*Main author for correspondence.

Abstract

This study investigates the mechanism by which knowledge spilled over from a firm's research publication consequently spills into the focal firm as a form of proprietary knowledge when it is engaged in an emerging science-related technology. We define the knowledge spillover pool (KSP) as an evolving group of papers citing a paper published by a firm. Focusing on the recent development of artificial intelligence, on which firms have published actively, we compare the KSP conditions related to the increase in patents created by the focal firm with those created by external actors. Using a Cox regression and subsequent contrast test, we find that both an increasing KSP and an increasing similarity between the idea published by the focal firm and KSP are positively related to the proprietary knowledge creation of both the focal firm and external actors, with such relations being significantly stronger for the focal firm than for external actors. On the contrary, an increasing proportion of industry papers in the KSP are positively associated with the proprietary knowledge creation not only by the focal firm but also by external actors to a similar degree. We contribute to the literature on selective revealing and to the firms' publishing strategies.

JEL classification: O34, O36

1. Introduction

Research publication by firms has long been of interest to scholars because it cannot be easily understood under the theoretical lens of neoclassical economics or from the perspective of strategic management. Despite decreasing publications by firms since the 1980s (Arora *et al.*, 2018), firms engaging in emerging science-related technologies¹ have continued to publish research

¹ The influence of technologies closely interacting with the realm of science has been increasingly emphasized in the contemporary knowledge-based society (Pavitt, 1987; Narin *et al.*, 1997; Van Looy *et al.*, 2007) although the degree of the interaction significantly varies across sectors (Ahmadpoor and Jones, 2017). The phenomenon has been captured by social scientists, using such terms as “science-related technology” (Freeman, 1997), “science-based technology” (Meyer-Krahmer and Schmoch, 1998), and “Mode 2 knowledge production” (Gibbons *et al.*, 1994). This study uses the term “science-related technology” to represent the mutual (rather than one-directional) relationship between science and technology.

papers (Alexy *et al.*, 2012; Simeth and Cincera, 2016; Anthes, 2017; Grassano *et al.*, 2019). When such technologies emerge, incumbent firms often invest in research and development (R&D) to seize the opportunities coming from these new technologies, with such investment resulting in research publications. From the resource-based view, disclosing a firm's knowledge through publication may harm its competitive advantage because it decreases the inimitability and uniqueness of its technology (Wernerfelt, 1984; Barney, 1991). Similarly, the private investment model asserts that revealing a firm's private knowledge reduces return on investment, which finally decreases the incentive to reveal its research outcomes publicly (Demsetz, 1967; Audretsch and Feldman, 1996). According to this strand of the literature, while firms' losses from publications are obvious, the return is relatively unclear (Jaffe, 1986; Kogut and Zander, 1992).

On the contrary, innovation management scholars have examined why firms might disclose their research outcomes to the public despite the existence of high risks. The motivations mentioned include attracting top researchers in the field and maintaining formal and informal linkages with academia (Rosenberg, 1990; Hicks, 1995), which may allow firms to absorb frontier knowledge in an emerging area (Cohen and Levinthal, 1990). Some scholars argue that firms publish when they fear the appearance of substitutes rather than imitation by competitors in the early stages of technological innovation (Polidoro and Toh, 2011). Another group of scholars has noted a defensive purpose of publishing, namely, to block anyone from enforcing patent rights in a knowledge area by forming prior art through publications (Bar-Gill and Parchomovsky, 2003; Johnson, 2014). Additionally, a firm that publishes can shape external knowledge and even change the competitive or collaborative behavior of external actors by intentionally letting rivals know its internal knowledge (Polidoro and Theeke, 2012; Alexy *et al.*, 2013).

However, while prior studies have discussed the motivational aspects of publishing by firms, the actual consequences of such firm behavior have been relatively ignored. Moreover, the expected negative and positive consequences of firms publishing research have usually been discussed separately. This study bridges these gaps in the literature, with a particular focus on the spill-in mechanism of knowledge spilled over from a firm's publishing. The concept of knowledge spillovers, while typically perceived negatively in the past, has recently been seen in the positive context of helping an originating firm vicariously learn from the external knowledge influenced by the knowledge spilled over from it (Yang *et al.*, 2010; Alexy *et al.*, 2013). In this vein, if a firm can create proprietary knowledge related to external researchers' reaction to the firm's publishing, this is one of the important positive consequences that the firm can enjoy through publishing. We define the knowledge spillover pool (KSP) as an evolving group of papers citing a published firm paper. Based on this definition, we compare the KSP conditions linked to the increase in proprietary knowledge creation (i.e., patents) by the originating firm (i.e., the firm that published a paper) with those of external actors.

In doing so, we focus on the recent development of artificial intelligence (AI) in 2006–2017 as an emerging science-related technology. Firms with assets complementary to AI have invested in relevant research and actively published their research outcomes over this period (see Figure B1 and Table B1). Given that various expected mechanisms could be related to the consequences of firms publishing in this field, we describe our empirical context in Section 2.3 and provide a rationale for focusing on the spill-in mechanism of proprietary knowledge represented by patents. Some properties of the KSP are expected to be linked to the creation of proprietary knowledge by the originating firm as well as external actors. In this study, research questions are derived based on the following three aspects: the size of the KSP, the proportion of industry papers in the KSP, and the similarity between the KSP and a published paper. We examine the extent to which each aspect is related to the creation of proprietary knowledge by the originating firm and external actors and how the degree of relatedness differs across these two groups. A Cox regression and subsequent contrast test are conducted using the data on publications and patent records by the firms engaged in basic research on AI.

This study makes several contributions to the innovation literature. We add to the ongoing discussion on the puzzling phenomenon of firms' publishing by providing empirical evidence on the consequences of such counterintuitive firm behavior. We link the insights from this study to the literature on selective revealing (Henkel, 2006; Alexy *et al.*, 2013) by discussing how firms engaging in the early stages of science-related technologies can ultimately create proprietary knowledge

connected to the knowledge spilled over from their research publications. Our results elaborate on how selective revealing through research publication could be another source of learning through the mechanism of the spill-in of spillover knowledge. We show the potential idiosyncratic advantages of reciprocating and when these advantages can be improved, thus linking the findings to the previous argument of the path-dependent evolutionary nature of technology development (Nelson and Winter, 1982). Moreover, this study suggests several managerial implications for firms that choose to selectively publish their research outcomes under uncertainty. We suggest how the focal firm can make search and prediction efforts to improve the effectiveness of its revealing strategies.

The remainder of this paper is structured as follows: Section 2 explains the key literature relevant to our research, points out the research gap, and develops a rationale for the focus of this study. Section 3 presents the research questions. Section 4 describes the empirical design of this research and data collection, while the findings are presented in Section 5. The study concludes in Section 6.

2. Conceptual and empirical background

2.1 Publishing as a strategy

Strategic management scholars have maintained that the acquisition of autonomous control over valuable and inimitable resources, including both tangible and intangible assets, is a key strategic dimension for a firm to build its competitive advantage (Barney, 1991). Teece (1986) argued that firms in sectors in which knowledge is a critical asset and easy to imitate should strive to legally protect their knowledge to benefit from appropriating knowledge. Therefore, publication by firms would seem to contradict the view of strategy scholars, as firms give up control over their knowledge and intentionally facilitate the outgoing knowledge spillover. Since published knowledge can be accessed by actors globally, the act of publication could thus promote the entrance of new firms into relevant markets (Jaffe, 1986). Given that some strategy scholars even assume that a source firm does not benefit when its knowledge spills over and is imitated by external actors (Kogut and Zander, 1992), publication by a firm appears to be counterintuitive behavior.

However, while publishing seems to be paradoxical for profit-seeking organizations, innovation scholars understand it as part of an open innovation strategy. The open innovation literature distinguishes two types of openness—inbound and outbound (Dahlander and Gann, 2010). The former refers to the sourcing or acquisition of the external resources required to innovate, whereas the latter involves the selling and revealing of internal resources. Within outbound innovation, a selling strategy offers direct pecuniary benefits to a source firm because the process includes the establishment and enforcement of intellectual property rights. By contrast, a revealing strategy does not result in direct monetary rewards to the source firm because it allows external actors to freely access its knowledge without facing legal exclusion (Henkel *et al.*, 2014). Although the appropriation regime that emerges after such revealing remains unclear and differs across fields, firms in various industries have often pursued a similar strategy. Interestingly, Allen (1984) found evidence of a firm's revealing strategy (referred to by the author as "collective invention") even in the early stages of iron production in 19th-century England. Firms in the iron industry actively shared production design rather than protecting it through patents.

Publication by a firm is within the scope of a revealing strategy, which is part of its outbound open innovation. Several motivations for firms' publishing behavior have been discussed, including attracting top researchers, maintaining links with academia (Rosenberg, 1990; Hicks, 1995), changing the competitive behavior of external actors (Polidoro and Theeke, 2012), and gaining legitimacy from the external environment (Nuvolari, 2004). Hayter and Link (2018) further found that the meanings and objectives of publishing by firms vary depending on whether the industry is in its emerging or maturity phase. In addition, Polidoro and Toh (2011) argued that firms decide to publish because they fear substitution, rather than imitation, in the early stage of developing a new technology.

Another interesting viewpoint on revealing behavior is that of Henkel (2006), who argued that firms strategically reveal part of their knowledge to induce collaborations with external

actors, given the difficulties arising from the uncertainties in partnering as well as the cost of finding and coordinating with appropriate and willing partners (Kale and Singh, 2009). This so-called “selective revealing” is characterized by the purposeful disclosure of selected resources to the public, including competitors. Alexy *et al.* (2013) showed that the mode of selective revealing can be problem- or solution-related, whereas the goal can be either path extension or path creation (Von Hippel, 1988). Alexy *et al.* (2013) categorized research publication as satisfying both solution-revealing and path-creating strategies, as the basic motivation is to create solution-related knowledge trajectories with the help of external actors, particularly when uncertainty is high and the appropriability regime is incomplete. Such an interpretation partially explains why incumbent firms have sought to publish their research outcomes during the emergence of science-related technology, which, by its very nature, is highly uncertain (Freeman, 1997).

In other words, recent arguments on selective revealing have emphasized that firms with internal knowledge can attract external actors who help improve relevant knowledge, thereby offsetting the cost of making, coordinating, and managing formal interorganizational relationships (Alexy *et al.*, 2013). In this vein, firms revealing their research outcomes through publications can attract academic researchers, who usually follow the norms of intrinsic motivation and peer group esteem (Dasgupta and David, 1994), and who can improve and validate the solutions suggested by firms. Such a possibility is in line with firms that selectively reveal software code to the public to leverage the capabilities of software developers, who are also known to be highly motivated by intrinsic factors and peer recognition (Henkel, 2006; Von Krogh *et al.*, 2012). Despite the risk of free riding, collaborative efforts produce new knowledge that might never appear if left to a single organization (Owen-Smith and Powell, 2004).

2.2 Knowledge spillover

Contrary to the substantial discussion on the motivation of firms’ publishing, relatively little attention has been paid to the consequences of a knowledge spillover from such revealing behavior. Knowledge spillovers are flows of knowledge from an investor who creates knowledge to external actors. Since knowledge is partially a public good characterized by non-rivalry and non-excludability, the spillover of knowledge occurs in many areas—even when the creator does not reveal such knowledge intentionally (Arrow, 1962). Therefore, economists have long discussed the difficulty of managing the tension between value creation and value appropriation when producing new knowledge, often finding that knowledge spillovers benefit society and external actors rather than the investor (Griliches, 1991; Romer, 1990).

Knowledge spillovers can, however, benefit investors through the “spill-in” mechanism. More specifically, Agarwal *et al.* (2007, 2010) described innovative activities as involving a dynamic process in which outgoing knowledge spillovers are verified and developed by external actors before eventually returning to the investor firm. In a similar vein, Yang *et al.* (2010) supported the existence of the spill-in mechanism by showing that certain characteristics of knowledge spillovers through patenting can positively influence the source firm’s gaining of future patents. However, the authors acknowledged that their study could not distinguish the positive aspects of spillovers from the negative ones.

To the best of our knowledge, despite the continuing debate on the meaning of a revealing strategy for a source firm, a balanced approach that considers both the positive and the negative aspects of the strategy’s consequences remains scarce in the literature. On the one hand, some scholars have highlighted the various motivations closely related to the potential benefits for a focal firm, while ignoring the loss typically incurred when knowledge flows to outside actors. On the other hand, some assume that spillovers and the resulting imitation usually impose a cost without offering any clear positive effects. This study bridges the gap by investigating the spill-in mechanism of an intentionally promoted knowledge spillover by firms publishing their basic research outcomes. As there are various possible dimensions of the consequences of knowledge spillovers in this field, we review the empirical context of AI and provide a rationale for our focus in the following section.

2.3 Corporate publishing in AI

The empirical context on which we focus is the recent development of AI (2006–2017). Although AI dates back to the 1950s, the technology has dramatically resurged recently with the advanced efficiency of hardware, huge amounts of data, and breakthroughs in learning algorithms. The methodological breakthrough is often credited to Hinton *et al.*'s (2006) work on the deep learning model (Taddy, 2018). The fundamental research questions addressed by AI researchers in both academia and industry concern machine improvement, which orients toward solving technological problems. The process of problem-solving largely interacts with the realm of basic science in terms of its reliance on science and the generation of new scientific knowledge (Rosenberg and Nelson, 1994; Denning, 2005; Boden, 2016), representing the science-related nature of AI technology.

Firms have noticed many potential opportunities emerging from AI, which also has features of general-purpose technology that can be applied to solve problems in various contexts (Teece, 2018). Therefore, some firms have started to invest in the basic R&D of AI and subsequently published a significant amount of research in the form of conference papers and journal articles. An important feature of the firms publishing on AI is their possession of big data as the key complementary assets required not only to perform AI research itself but also to capture the value from the relevant innovation. Hartmann and Henkel (2020) argued that firms engage in and publish AI research because they can benefit most from the advancement in the field due to their possession of the data.

In addition to possessing the data, another crucial complementary capability is required to benefit from the advancement of AI. Given the general purpose of AI, the key capability to take advantage of the relevant frontier knowledge is customizing and embedding upstream AI research outcomes to solve problems in downstream application domains (Boden, 2016; Norton, 2016). In other words, to implement a function in a product or service based on AI, the relevant AI algorithms obtained from basic research should be customized and embedded in the particular context of the application. These downstream technological activities are directly related to firms' profits.

A significant proportion of knowledge related to customizing and embedding AI algorithms as well as the AI algorithms themselves can be accumulated as firm-specific proprietary knowledge. To maintain such knowledge as proprietary, a firm can choose to use informal methods of protection such as trade secrets and complex designs. When imitation and reverse engineering are relatively easy, firms rely on formal intellectual property rights including patents although debates on the usefulness of formal and informal protection methods are ongoing (Cohen *et al.*, 2000; Arundel, 2001; Somaya, 2012; Granstrand, 2018; Foss-Solbrekk, 2021²). For example, IBM publishes a large amount of AI research but, at the same time, has a large AI-related patent portfolio and does not open key knowledge underlying the IBM Watson platform.³

Based on this background, among the various consequences potentially arising from a firm's publishing, this study focuses on the creation of proprietary knowledge represented by patents linked to the knowledge spilled over from the focal firm's publishing. Owing to the aforementioned nature of AI technology, such patents can encompass both downstream and upstream technological activities related to AI. Original technological ideas related to AI that satisfy patentability requirements such as novelty, non-obviousness, and usefulness can be patented (Foss-Solbrekk, 2021), while software and source code that implement such ideas (i.e., expression of ideas, but not the ideas themselves; the expression is what copyright protects) are likely to be protected through copyright-based proprietary licenses if not freely released.⁴ Recent statistics show that a considerable number of AI-related patents have been increasingly applied for over the past decade (WIPO Technology Trends, 2019; USPTO IP Data Highlights, 2020).

² Foss-Solbrekk (2021) argued that broadening the scope of patent protection for AI technologies is important as well as reported the recent trend in this direction (see the European Patent Office Guidelines for Examination 3.3:1 Artificial Intelligence and Machine Learning). Many patent offices consider inventions related to AI technology as patentable matter despite examination criteria differing across countries to some extent.

³ <https://www.ibm.com/services/artificialintelligence>.

⁴ Several popular AI software frameworks (e.g., TensorFlow and PyTorch) have been freely released by large technology companies under free software licenses.

Although patents cannot perfectly capture the whole range of proprietary technological knowledge, many studies have used patents as a proxy for firms' proprietary technological knowledge for two main reasons. First, patenting is costly, implying that patented knowledge is likely to include a valuable part of a firm's knowledge (Basberg, 1987). Second, knowledge published by a firm (through patents or papers) is closely connected to the unpublished part of the firm's knowledge stock (Hicks, 1995), meaning that patents can represent a considerable extent of a firm's proprietary knowledge base.⁵

Knowledge spillovers happen when external actors use the knowledge flowing from an originating firm to create their own knowledge (Griliches, 1991). Accordingly, the use of revealed knowledge by external actors creates a pool of knowledge that spills over from an originating firm. In this study, we define a KSP as a group of papers that has cited a paper published by an originating firm. The KSP represents an extended and validated set of knowledge that reflects collaborative inputs from various researchers in the relevant area. This definition partly follows Yang *et al.* (2010), who defined a firm-level KSP as a group of patents including not only the patents citing a focal firm's patents but also complementary patents recombined with the focal firm's patents to create new patents, which draws on the evolutionary concept of recombinatorial search (e.g., Fleming, 2001). Our definition differs from that of Yang *et al.* (2010) in two ways. First, the KSP is defined at the published paper level. Second, the KSP does not include knowledge recombined with revealed knowledge because this study focuses on the conditions of an evolving body of knowledge created after the focal firm has published. Based on the concept of the KSP, we investigate and compare its conditions that are related to the creation of proprietary knowledge represented by the patents of an originating firm and external actors.

Our definition of the KSP corresponds to a set of knowledge advanced through the help of researchers⁶ after an originating firm publishes its research. Therefore, the mechanism we address reflects conceptual discussions on selective revealing, namely, that firms reveal part of their knowledge to encourage external actors to collaborate to create and advance relevant new knowledge, which corresponds to our KSP, and then try to exploit the advanced set of knowledge (Alexy *et al.*, 2013). Similarly, the KSP is also in line with discussions on the production of Mode 2 knowledge (Gibbons *et al.*, 1994). These scholars argue that the central challenge of developing an emerging technology that requires the participation of various organizations is how to co-create a pool of knowledge, which corresponds to our KSP, and how to subsequently take advantage of the co-produced set of knowledge. Our focus is also rooted in the underlying nature of technological innovation, which is a cumulative and path-dependent process proceeded by knowledge spillovers among numerous actors (Nelson and Winter, 1982). Lastly, the operationalization draws on bibliographic evidence that shows frequent direct citation linkages between papers and patents in science-related technologies such as biotechnology and AI (Murray, 2002; Ahmadpoor and Jones, 2017).

3. Research questions

We derive the following three main research questions on the characteristics of an evolving KSP that can be related to the creation of proprietary knowledge by a focal firm and external actors. The main aspects on which we focus are the size of the KSP, proportion of industry papers in the KSP, and similarity between the KSP and a published paper.

First, the size of the KSP is the extent to which a paper published by a focal firm is cited by subsequent papers that form a set of new knowledge. For a firm to learn from external actors, its knowledge stock should be connected to that of external actors because such relevance reduces search costs and expedites knowledge recombination (Cohen and Levinthal, 1990). The KSP provides an efficient source of search for an originating firm because, from the above definition of the KSP, every piece of knowledge in the pool has potential relevance to the originating firm's

⁵ The defensive purpose of patenting is relatively prevalent in the emerging period of AI. However, patenting inherently has mixed purposes including both offensive and defensive roles (Granstrand, 1999), which is also supported by the fact that there are many other routes of making prior art without spending as much money as patenting. We deem firms' patents with those mixed purposes as part of proprietary knowledge.

⁶ Researchers here indicate individuals who contribute to knowledge production.

idea. Hence, as the size of the KSP increases, it is expected that an originating firm is provided with better conditions to vicariously learn from the evolving KSP.

Moreover, a large KSP can increase the possibility that the focal paper has contributed to the formation of the knowledge trajectory and the direction of knowledge evolution within a particular paradigm (Abernathy and Utterback, 1978; Dosi, 1982). During the emergence of a new technology, the relevant knowledge trajectories remain unclear and extreme uncertainties exist in every dimension. Relative stability can only be achieved once the trajectories are determined, implying that a significant number of actors have committed to support a certain technological direction (Garud and Rappa, 1994; Grant, 1996). Such a collective commitment results in the legitimization of uncertain knowledge, eventually creating a new knowledge-intensive industry (Garud *et al.*, 2002). Considering such characteristics of an emerging technology, a firm's influence on the formation of an emerging knowledge trajectory through publishing can be regarded as a successful informal collaboration with researchers outside the firm (Alexy *et al.*, 2013). This is because, by doing so, it is likely that the originating firm can induce knowledge evolution to leverage its resources. Therefore, when the size of the KSP increases, a focal firm's influence on the knowledge trajectory can rise, making it more likely to find useful technological opportunities from the KSP.

By contrast, a large KSP also implies that external actors may have found opportunities to expand their technological knowledge base using the knowledge revealed by an originating firm. A large KSP can be a strong indication of the existence of numerous external actors, including competitors, which can exploit the originating firm's knowledge (Grant, 1996). This implies that the potential hazard from such revealing might be extensive, especially in terms of competitors' entrance into the relevant field. Furthermore, as described above, if a paper creates a sizable KSP, this increases the possibility that the revealed solution has become part of a knowledge trajectory. The existence of common knowledge facilitates understanding and communication among heterogeneous actors (Zucker, 1987), ultimately allowing them to achieve efficient knowledge integration using the common knowledge base. Therefore, as the size of the KSP increases, the possibility that numerous actors can efficiently create proprietary knowledge by drawing on the relevant KSP also increases. Subsequently, we examine when the focal firm's proprietary knowledge creation increases over that of external actors in relation to the size of the KSP. This brings out the following research questions.

RQ1: Is the size of the KSP formed by a firm's publication positively related to the (i) *originating firm's* and (ii) *external actors'* patent application linked to the KSP? (iii) If so, to what extent does the relation between the two groups differ?

The second facet is the proportion of industry papers in the KSP, which represents how many papers within the KSP are from industry. Following Arora *et al.* (2018), an industry paper here is defined as one in which at least one of the authors is affiliated with a firm. A high proportion of industry papers within the KSP imply that most actors interested in the focal firm's research topic are commercially oriented. Although the number of university researchers who seek to create economic value using their research outcomes has increased (Etzkowitz, 2003), firm researchers are more directly linked to commercial activities than university researchers. In this vein, a KSP consisting of a high proportion of industry papers is likely to include ideas closely related to patentable knowledge, which are more likely to be used as a means to gain commercial advantage.

At the same time, a high proportion of industry papers in the KSP imply that many industry actors are interested in the technology area, which suggests a strong competition among firms seeking to preempt relevant patent applications. This intensified competition increases the possibility that rival firms apply for the key patents linked to the KSP even before the originating firm gains such an opportunity. Moreover, during technology emergence, firms have a strong incentive to lead the paradigm of technology evolution in a direction favorable to their own businesses and resources. Hence, to create lock-in effects, which make subsequent inventions created in a path-dependent manner, competing firms are likely to seek patenting in relevant technology areas as early as possible. Furthermore, when there is fierce competition over an uncertain emerging technology, firms are also likely to pursue patenting for defensive purposes, namely, to block other actors from enforcing patent rights (Granstrand, 1999). Subsequently, we examine whether the

focal firm's patent application connected to the KSP increases over that of external actors as the proportion of industry papers in the KSP increases. We derive the following research questions:

RQ2: Is the proportion of industry papers in the KSP formed by a firm's publication positively related to the (i) *originating firm's* and (ii) *external actors'* patent application linked to the KSP? (iii) If so, to what extent does this relation between the two groups differ?

The last dimension is the similarity between the knowledge revealed by a focal firm and the relevant KSP. Here, "similarity" refers to how objectively and accurately the KSP created by external actors addresses the specific interests of an originating firm (Cohen and Levinthal, 1990; Lane and Lubatkin, 1998). For firms to work on an external actor's solution, great efforts are needed to match the solution to their own knowledge structure (Kotha *et al.*, 2013). The similarity between the revealed solution and KSP can be gradually constructed through these efforts to assimilate, which involves a tacit process accompanied by high costs of translation.

Therefore, when a firm is planning to engage in formal R&D collaboration, one of the most challenging tasks is the identification of a suitable partner who can reduce the assimilation cost and create a complementary synergy (Kale and Singh, 2009). In the emerging stage of a science-related technology, publishing an uncertain and incomplete research outcome can be an efficient way for firms to reduce the costs of finding a suitable collaboration partner as long as external actors provide voluntary support for the cost of assimilation (Alexy *et al.*, 2013). Therefore, as the similarity between the revealed knowledge (i.e., a publication) and relevant KSP rises, it is expected that quasi-collaboration through a revealing strategy can be closely related to the validation and enhancement of the solution addressing the needs of the originating firm. Under high levels of similarity, originating firms can efficiently monitor and comprehend the information within the KSP, and therefore, they are expected to more easily create new proprietary knowledge by absorbing external information.

However, at the same time, R&D collaboration is a bilateral interaction between (or among) actors rather than a one-directional relationship from one to another (Katz and Martin, 1997). Therefore, the higher the similarity between the revealed solution and relevant KSP, the more likely it is that external actors have an interest in the solution published by the originating firm. Those external actors may have considerable capability to interpret and exploit the knowledge within the KSP. In this vein, Katz and Martin (1997) argued that collaboration is meaningful only when there exists a clear division of labor between (or among) actors because a highly overlapping background is likely to result in redundant outcomes as well as more intensive competition for capturing appropriable outputs. For a similar reason, the alliance literature has also highlighted that a complementary rather than a similar resource background is a fundamental condition for the successful cooperation (Mowery *et al.*, 1996; Dyer and Singh, 1998; Makri *et al.*, 2010).

In addition, an increasing amount of similar technological knowledge can decrease the uncertainty surrounding a technology (Polidoro and Theeke, 2012), encouraging further participation by actors in that R&D area. This is because investment by many actors with similar knowledge implies not only a confidence in the validity of that type of technology (Garud *et al.*, 2002) but also a significant level of refinement that might have strengthened the specific technological knowledge (Anderson and Tushman, 1990). Subsequently, we examine whether the focal firm's patent application connected to the KSP increases over that of external actors as the similarity between the knowledge revealed by a firm and the corresponding KSP increases. We arrive at the following research questions:

RQ3: Is the similarity between the knowledge revealed by a firm and corresponding KSP positively related to the (i) *originating firm's* and (ii) *external actors'* patent application linked to the KSP? (iii) If so, to what extent does this relation between the two groups differ?

4. Methodology

4.1 Data

To answer the above research questions, we collect a list of top-tier conferences and journals from Guide2Research, a portal providing the rank of journals, conferences, and authors in computer science. Among several areas of computer science, the subfield titled "Machine Learning, Data Mining, and Artificial Intelligence" is chosen for this study. From this subfield, the top 10

conferences and journals are listed (Appendix A). Then, we collect information on the 56,981 papers published by the selected journals and conferences during 2006–2016 from the SCOPUS database. In computer science, top-tier conferences have a higher status than most journals, excluding a few highly ranked ones (Vardi, 2009; Freyne *et al.*, 2010). Firms use conferences to advertise research areas of their interest to talented researchers, particularly by sponsoring top conferences. Moreover, firms' investment in computer science research substantially correlates with their conference sponsorship decisions and research publications in the field (Baruffaldi and Poege, 2020). Therefore, we select the publicly listed firms that sponsored one of the listed conferences, and among them, we obtain 86 firms that published at least one article in the field during 2006–2016. These firms cover the major industry actors in AI that appeared in the recent articles such as World Intellectual Property Organization (WIPO Technology Trends, 2019) and Hartmann and Henkel (2020). We follow Arora *et al.* (2018) to define a firm paper as one in which at least one of the authors is firm-affiliated. From 2006 to 2016, 5835 papers were published by the chosen firms. Figure B2 in Appendix B shows the number of firms publishing in the selected conferences and journals by country. Most of the firm papers were published by firms in the United States, followed by Japan and China. Table B1 in Appendix B shows the number of papers published in the venues listed by major firms, including Microsoft, Google, and IBM. In addition, Table B2 in Appendix B reports the number of publications by major organizations, excluding firms. It shows the leading role of US and Chinese universities, followed by Japanese and some European institutions. Overall, the number of papers published by major firms in AI during this period was comparable to that by leading universities.

To calculate the variables of interest, the per-paper KSP is formed for the 5835 firm papers using the citation linkages among all the 56,981 papers collected in this field.⁷ Since the KSP of a paper in this study indicates a group of papers citing the focal paper, the KSP evolves as the number of papers that cite the focal paper increases (or at least remains the same) over time. This study considers the state of the per-paper KSP annually because the timing of paper publication is recorded by year. This study is interested in the conditions under which a paper's KSP is likely to be linked to the patent application of the revealing firm or external actors. To make such linkages, we use the PATSTAT database, which includes worldwide patent information collected from patent offices globally. Using the non-patent reference information in this database,⁸ citation linkages between the within-KSP papers and subsequent patents are captured.⁹

4.2 Variables and methods

4.2.1 Dependent variables and models

The research questions of this study address the conditions under which the rate of patent application linked to the KSP of a firm paper is increased by two groups: the focal firm and external actors. Therefore, the dependent variable should reflect the rate of patent application, which cites any papers within each KSP, by the focal firm and external actors. In addition, the independent variables of interest address the characteristics of the KSP that evolve over time. Following prior studies (Podolny and Stuart, 1995; Nerkar and Paruchuri, 2005; Marco, 2007), we use a recurrent event hazard rate analysis to model the rate of patent application. Cox regression can be extended to model the hazard rate of repeated events as well as reflect time-varying covariates (Cox, 1972; Kalbfleisch and Prentice, 1980; Cook and Lawless, 2007). The equation takes the following specification:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta z_i + \gamma x_i(t)),$$

⁷ Given the general-purpose nature of AI, a focal firm's paper can be cited by papers in a variety of domains that apply the suggested methods. Although the application of AI algorithms can be linked to a valuable spill-in for the focal firm, setting the boundary of the KSP as outlets in all potential fields of study makes it difficult to maintain knowledge quality in the KSP. Therefore, we consider the KSP within papers published in the highly reputed outlets in the field of AI, focusing on the extended and validated version (even minor) of the AI algorithms suggested by the focal firm.

⁸ Previous studies showed that the proportion of examiner citations is ignorable in the case of non-patent citations (Lemley and Sampat, 2012; Ahmadpoor and Jones, 2017; Bikard and Marx, 2020). In line with this, our data show that examiner-added citations account for only about 1% of all citations.

⁹ Given that the list of references can be revised over the lifetime of a patent, this study uses an updated version of the reference information available in January 2018.

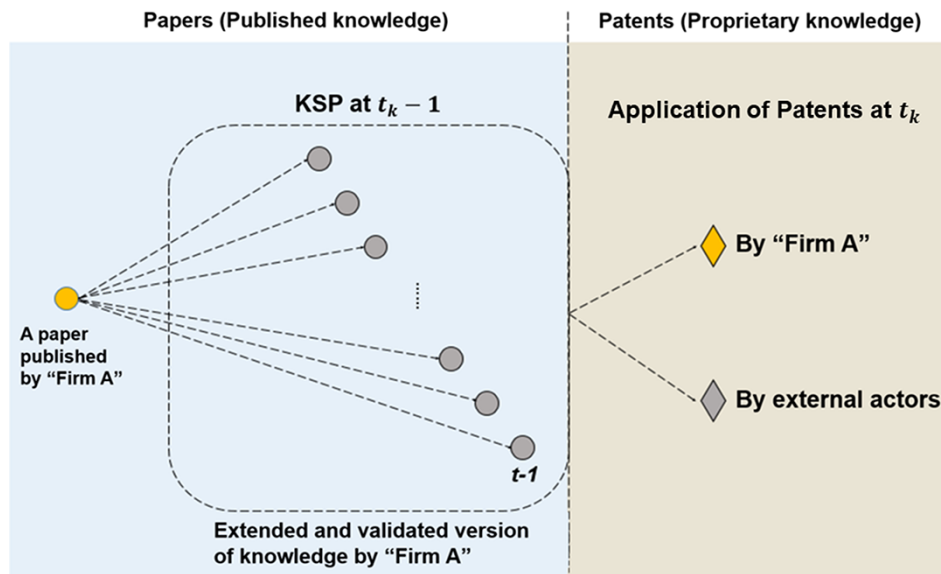


Figure 1. Research design

where $\lambda_i(t)$ is the application rate of patents that cite one of the papers in the KSP of paper i from time t to $t + dt$; $\lambda_0(t)$ is a baseline citation rate that does not make any assumption about its distribution; z_i indicates the vector of time-invariant covariates for the static properties of a firm paper; and $x_i(t)$ is the vector of the time-varying covariates. The time gap between patent applications that cite any paper within each per-paper KSP is used as a dependent variable. In other words, the time from the year of a paper's publication to the year of the first patent application connected to a particular KSP is regarded as the first event; then, the times between subsequent patent application events are used sequentially. To obtain these values, paper–patent citation linkages are constructed by relying on the titles of the papers cited in each patent. Figure 1 describes the research design.

In addition, the linkages are divided into two groups depending on whether the applicants of the patent include the focal firm: one group of events is for the focal firm's patent application and another group concerns external actors' patent application. Based on the events linked to each KSP, the time gap in patent application is computed for these two groups. The model reflects the censored time for every per-paper KSP as 2017. For the focal firm's patent filings related to the KSP, there are 146 events. In addition, for each of the 5835 papers, censoring at the end of the observation period must be included, yielding a total of 5981 records. For external actors' patent filings related to the KSP, there are 5333 events. For each of the 5835 papers, censoring needs to be included, yielding a total of 11,368 records. Lastly, as there are multiple observations for each per-paper KSP, we use robust standard errors clustered by paper in the estimation.

4.2.2 Independent variables

Our research questions include three time-varying covariates related to the evolving KSP: the size of the KSP, firm proportion in the KSP, and similarity between a paper published by the firm and corresponding KSP. We choose a 1-year time lag between the patent application event and state of the KSP reflected in the regression model (see Figure 1). Specifically, for the creation of a patent citing any papers within a KSP in year t_k , we reflect the KSP of year $t_k - 1$ in the regression model. This time lag is determined following Yang *et al.* (2010).

4.2.2.1. Size of the KSP

First, the size of the KSP is measured using the number of papers that cite the original publication of a focal firm. This is a time-varying variable that increases (or at least does not decrease) over

time because the number of forward citations of a paper is cumulated. Suppose paper i , which was published by a firm in 2006. If one of the papers in the KSP of paper i is cited by a patent applied in 2011 (t_k), the size of the KSP corresponding to this patent application event is measured in 2010 ($t_k - 1$). In this case, the size of the KSP is considered as the number of papers that cited paper i from 2006 to 2010.

4.2.2.2. Proportion of industry papers in the KSP

Second, the proportion of industry papers in the KSP is calculated using the author affiliation information. As defined, a firm paper has at least one author affiliated to the firm. This variable is a time-varying one because the proportion of industry papers within each KSP changes as the size of the KSP increases over time. If a patent that cites one of the papers within the KSP of paper i is created in 2011 (t_k), the proportion of industry papers in the KSP corresponding to this patent application event is measured in 2010 ($t_k - 1$). Specifically, the proportion of industry papers in this KSP is considered as the number of firm papers within the KSP divided by the number of papers that cited paper i from 2006 to 2010.

4.2.2.3. Similarity between the focal paper and KSP

Third, the similarity between the firm paper and KSP is calculated by depending on the text similarity between the two groups of documents. To do this, we use the dynamic topic model (DTM), a topic modeling technique used to infer the latent topic underlying a collection of documents and per-document topic proportions (Blei and Lafferty, 2006). The DTM is different from static models in that it reflects the evolution of the words composing each topic over time. Because we analyze papers published over 11 years, it is difficult to expect topics extracted from these documents to consist of consistent words over this period. Therefore, the use of the DTM is reasonable to capture the evolution of topics in a more flexible way. Based on the 56,981 papers published during 2006–2016, the DTM is formulated to capture the topics representing those documents and per-paper topic proportions. The keyword information of each paper is used for the text analysis because it includes what the paper is about rather than involving overly broad and general terminologies.

In addition, the number of topics should be determined for the modeling. Although scholars have suggested the best fitting model for finding the optimal number of topics, such methods have been criticized because they often produce too many topics that do not represent distinct meanings (Chang *et al.*, 2009). Instead, constraining the number of topics, typically to around 100, provides more meaningful results (Blei and Lafferty, 2007; Hall *et al.*, 2008). Recently, scholars have also adopted this finding (Kaplan and Vakili, 2015). Following these arguments, we constrain the model to 100 topics.

By conducting the topic modeling, we obtain per-paper topic proportion vectors for the 56,981 papers. Based on this information, the vectors corresponding to the papers within each KSP are averaged to allow us to compute the cosine similarity between each paper published by the firm and its corresponding KSP. Since the KSP evolves over time, the similarity also changes over time. If a patent that cites one of the papers within the KSP of paper i is created in 2011 (t_k), the similarity is measured in 2010 ($t_k - 1$). The similarity between the focal paper i and corresponding KSP for a patent application event in year t_k is as follows:

$$Sim_{i, t_k} = \frac{A \cdot B}{\|A\| \|B\|},$$

where A and B are the vectors of the focal paper and corresponding KSP in year $t_k - 1$, respectively.

4.2.2.4. Group dummies

To examine the significance of the variables of interest in each group as well as further test the difference in the variable effects between the groups, *Group dummy 1* and *Group dummy 2* are used. These variables have external actors and the focal firm as the reference group, respectively. We add the interaction terms between each explanatory variable and *Group dummy 1* to examine

the significance of the variables of interest in the focal firm group. Similarly, the interaction terms between each explanatory variable and *Group dummy 2* are added to examine the significance of the variables of interest in the external actors group. Lastly, a single term of *Group dummy 1* is added to the model to capture the difference in the hazard rate when the variables are not reflected.

4.2.3. Controls

We add control variables that are expected to be related to the occurrence of patent application events. First, the diffusion of knowledge is generally known to follow an S-curve. Following previous studies employing a recurrent event hazard rate modeling approach (Podolny and Stuart, 1995; Nerkar and Paruchuri, 2005; Marco, 2007; Jee *et al.*, 2019), we add the time gap between the publication of the firm paper and creation of the patent citing that paper's KSP as well as its squared term as controls (*Years from publication*, *Years from publication squared*). We include the number of author affiliations in each firm paper because the diversity of author affiliations positively affects the influence of the paper (Franceschet and Costantini, 2010) (*Number of author affiliations*). We control for the number of authors to reflect whether papers with multiple authors are likely to be more influential than single-authored papers (Gazni and Didegah, 2011; Didegah and Thelwall, 2013) (*Number of authors*). We control for the number of references in each firm paper (*Number of references*), given that this reflects the field-wide potential of it to influence future knowledge production (Garfield, 1979). Well-known scholars are likely gain credit from future works by others (Merton, 1968). Hence, this may influence their ultimate contribution to other future works, including patent applications. To capture this, we control for the maximum performance among the authors of each focal paper. Individual author performance is measured using the number of papers published by the author in top-tier conferences and journals in the 5 years before paper publication (*Max author performance*). To address unobserved heterogeneity in our research design, we control for the number of patent application events that occurred on a particular per-paper KSP before the patent application event in year t_k divided by the age of the focal paper (*Number of prior events divided by age of focal paper*) (Heckman and Borjas, 1980). As indicated by prior studies (Podolny and Stuart, 1995; Nerkar and Paruchuri, 2005; Marco, 2007; Jee *et al.*, 2019), this variable can control for the time-constant effects of unobserved factors that produce variance in each KSP's disposition to be cited by a patent. Since it is expected that firm-level intensity of technological activity is positively related to the originating firm's patent application connected to the KSP, we control for the total number of patent applications in year t_k by the originating firm (*Number of patents by the originating firm*). In addition, we control for *Firm size* based on the total assets (in millions of US dollars) of each firm in year t_k . Lastly, we also control for the industry dummy using two-digit North American Industry Classification System (NAICS) codes (*Industry dummy*). The Compustat database is used to obtain the yearly total assets and sector information of the publicly listed firms in our sample.

5. Findings

Table 1 contains the basic statistics and correlations. The industry dummy shows that firms are dispersed across 13 sectors, with the majority concentrated in two broad categories: manufacturing and information. Table 2 shows the results of the Cox regression with respect to patent application by the focal firm and external actors. Table 3 presents the results of the contrast test that examines whether the differences between the focal firm and external actors are significant in terms of the coefficients of the variables of interest.

Model 1 in Table 2 includes only the control variables. First, *Years from publication squared* is negatively significant, implying that the rate of patent application connected to the KSP of a paper generally shows an inverted U-shaped pattern following the year of a firm's paper publication. *Group dummy 1*, which has external actors as its reference group, is negatively significant, showing that the default hazard rate of external actors' patent application is higher than that of the focal firm. The maximum performance of the authors of a paper (*Max author performance*) is positively significant. This implies that when a firm publishes a paper written by a research team, including a high-performing researcher, the KSP of the paper is likely to be linked to more patent

Table 1. Descriptive statistics and correlations

	Mean	S.D.	Min	Max	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)
1) Size of the KSP (in year $t_k - 1$)	11.40	25.7566	0	301	-										
2) Proportion of industry papers in the KSP (in year $t_k - 1$)	0.21	0.2870	0	1	0.0679	-									
3) Similarity between the focal paper and KSP (in year $t_k - 1$)	0.34	0.3132	0	1	0.2588	0.3918	-								
4) Years from publication (in year t_k)	4.95	2.8572	0	11	0.1343	0.0516	0.1175	-							
5) Number of author affiliations	2.28	1.0494	1	24	0.0333	-0.0971	-0.0304	-0.0785	-						
6) Number of authors	3.82	1.9609	1	69	0.0005	-0.0070	-0.0114	-0.1083	0.4303	-					
7) Number of references	27.52	14.0487	1	239	0.1008	-0.0553	0.0135	-0.2089	0.1504	0.0530	-				
8) Maximum performance of authors in the focal paper	6.63	6.4244	1	45	0.0340	0.0596	0.1173	-0.1325	0.1554	0.1144	0.0801	-			
9) Number of prior events/age of the focal paper	0.62	1.5752	0	12.2	0.8086	0.0663	0.2342	0.0386	0.0267	0.0154	0.0827	0.0282	-		
10) Number of patents by the originating firm (in year t_k)	3867	4871	0	56,580	0.3307	0.1024	0.2014	-0.0765	-0.0113	-0.0567	-0.0039	0.0659	0.4322	-	
11) Firm size (in year t_k)	135,067	95,698	0	781,818	0.0598	-0.0264	0.0024	0.1190	0.0119	-0.0324	0.0110	0.0758	0.0168	0.2191	
12) Industry dummy	Two-digit NAICS (among the 86 firms, 40 are in manufacturing [NAICS 33], 30 firms are in information [NAICS 51], and the remaining 16 firms are dispersed across 11 other sectors).														
13) Group dummy 1	Focal firm versus external actors (reference group), external actors: 11,368 (events: 5533), focal firm: 5981 (events: 146).														
14) Group dummy 2	External actors versus focal firm (reference group).														

Table 2. Results of the Cox regression

	Model 1	Model 2	Model 3	Model 4	Model 5
Size of the KSP [*] Group dummy 1 (focal firm)		0.0229 ^{***} (0.0022)			0.0228 ^{***} (0.0018)
Size of the KSP [*] Group dummy 2 (external actors)		0.0068 ^{***} (0.0017)			0.0068 ^{***} (0.0014)
Proportion of industry papers in the KSP [*] Group dummy 1 (focal firm)			1.0131 ^{***} (0.1412)		0.2425 (0.2053)
Proportion of industry papers in the KSP [*] Group dummy 2 (external actors)			0.4469 ^{***} (0.0901)		0.1827 [*] (0.0972)
Similarity between the focal paper and KSP [*] Group dummy 1 (focal firm)				2.5272 ^{***} (0.2354)	2.2838 ^{***} (0.2691)
Similarity between the focal paper and KSP [*] Group dummy 2 (external actors)				1.1323 ^{***} (0.0943)	1.0808 ^{***} (0.0955)
Group dummy 1 (focal firm)	-2.5377 ^{***} (0.1153)	-2.6760 ^{***} (0.1123)	-2.6609 ^{***} (0.1233)	-3.0532 ^{***} (0.1679)	-3.1532 ^{***} (0.1736)
Years from publication	0.0557 [*] (0.0295)	0.0294 (0.0307)	0.0415 (0.0296)	-0.0469 (0.0294)	-0.0714 ^{**} (0.0301)
Years from publication squared	-0.0212 ^{***} (0.0029)	-0.0201 ^{***} (0.0030)	-0.0199 ^{***} (0.0029)	-0.0127 ^{***} (0.0027)	-0.0118 ^{***} (0.0027)
Number of author affiliations	-0.0777 ^{**} (0.0391)	-0.0824 [*] (0.0380)	-0.0656 [*] (0.0384)	-0.0774 (0.0373)	-0.0728 ^{**} (0.0368)
Number of authors	-0.0299 (0.0174)	-0.0296 (0.0215)	-0.0293 (0.0208)	-0.0327 (0.0207)	-0.0342 [*] (0.0201)
	0.0023 (0.0021)	0.0009 (0.0021)	0.0031 (0.0020)	0.0016 (0.0022)	0.0007 (0.0021)
Maximum author performance	0.0150 ^{***} (0.0042)	0.0155 ^{***} (0.0040)	0.0141 ^{***} (0.0043)	0.0097 ^{**} (0.0045)	0.0104 ^{**} (0.0043)
Number of prior events divided by the age of the focal paper	0.2587 ^{***} (0.0280)	0.1873 ^{***} (0.0379)	0.2617 ^{***} (0.0274)	0.2591 ^{***} (0.0235)	0.1888 ^{***} (0.0300)
Number of patents by the originating firm	0.0001 ^{***} (9.07e-6)	0.0001 ^{***} (9.16e-6)	0.0001 ^{***} (9.24e-6)	0.0001 ^{***} (9.06e-6)	0.0001 ^{***} (9.16e-6)
Firm size	-4.05e-6 ^{***} (6.44e-7)	-4.44e-6 ^{***} (6.25e-7)	-3.90e-6 ^{***} (6.40e-7)	-3.92e-6 ^{***} (6.44e-7)	-4.24e-6 ^{***} (6.26e-7)
Industry dummy	Yes	Yes	Yes	Yes	Yes
-2 Log-likelihood	97,459.654	97,255.025	97,349.806	96,822.869	96,628.75

Values in parentheses are robust standard errors clustered by firm papers.

Group dummy 1 (focal firm vs. external actors): external actor is the reference group.

Group dummy 2 (external actors vs. focal firm): focal firm is the reference group.

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$.

application. For the *number of authors* and *number of author affiliations*, the model shows negatively significant and insignificant coefficients, respectively. Such results are not counterintuitive in light of the considerable field-level difference in the degree of attention received by multiple- and single-authored papers (Smart and Bayer, 1986; Bridgstock, 1991) because the papers we analyze cover several subfields of AI. The coefficient of the *Number of references* is insignificant, differing from Garfield's (1979) expectation of a positive relation between the number of references in a paper and its impact. The *Number of prior events divided by age of the focal paper* shows a

Table 3. Contrast test between the focal firm and external actors

	Difference in coefficients [focal firm – external actors]
Size of the KSP	0.0160*** (0.0023)
Proportion of industry papers in the KSP	0.0598 (0.2266)
Similarity between the focal paper and KSP	1.2030*** (0.2847)

Values in parentheses are standard errors. The contrast test is based on the full model (i.e., model 5 in Table 2).

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$.

positive and significant result. This outcome supports the existence of the time-constant effects of unobserved factors resulting in variance in each KSP's disposition to be cited by a patent. The *Number of patents by the originating firm*, a firm-level control for the intensity of technological activity of the originating firm, shows a positive and significant result as expected. Lastly, the coefficient of *Firm size* is negatively significant, implying that the KSP of publication by larger firms is less likely to be associated with the creation of patents.

We now turn to the regression results. First, model 2 shows a positively significant relation between the size of the KSP and rate of relevant patent application by both the focal firm and external actors. However, the results of the contrast test in Table 3 show that the effect of the size of the KSP on patent application is significantly larger in the focal firm group than in the external actors group. In other words, the increasing KSP of a paper published by a focal firm provides a more useful pool for searching for new technological opportunities for the focal firm compared with external actors. In summary, from the perspective of the gain in proprietary knowledge, the focal firm's gain from publication rises when its publication is cited by future studies and more so than the gain of external actors. In this vein, if a firm understands the conditions under which its publication can receive more attention from scholars, selective publication satisfying these conditions can be a more useful strategy for gaining greater proprietary knowledge.

Second, in model 3, the proportion of industry papers in the KSP is positively significant in both the focal firm and the external actors group.¹⁰ That is, the proportion of industry papers within the KSP is positively linked to the rate of patent application affected by the KSP for both the focal firm and external actors. In addition, the results in Table 3 show no statistical difference between the two groups in terms of the significance of the proportion of industry papers within the KSP. In summary, the higher proportion of industry papers in the KSP states that the published idea is more likely to be connected to the focal firm's proprietary knowledge creation; however, at the same time, many potential market competitors have the capability to understand and utilize the technology. Considering the results, when a KSP evolves toward showing a high proportion of industry papers within it, the firm may take advantage only when it has the complementary capabilities and resources needed to capture value from the knowledge as quickly as possible.

Third, model 4 shows that the similarity between the knowledge within a published paper and KSP is positively significant for both the focal firm and external actors. In other words, as the KSP is formed around knowledge similar to the published idea, the rate of patent application connected to the KSP increases for the focal firm and external actors. However, the results in Table 3 show that the coefficient of the similarity between the focal paper and KSP is significantly higher for the focal firm than for external actors. Therefore, when the KSP evolves toward being highly similar to the idea published by the focal firm, this direction of the KSP evolution leads to greater proprietary knowledge for it compared with external actors.

¹⁰ The relative significance of the coefficient of the proportion of industry papers in the KSP weakens in the full model (model 5).

6. Discussion and conclusion

Research outcomes are extensively published by firms although whether such behavior delivers positive consequences to them remains uncertain. Based on the concept of knowledge spillovers, this study sheds light on the mechanism underlying the spilling in of the revealed knowledge of firms investing in R&D on emerging science-related technology. In particular, we investigate the conditions under which the knowledge published by the firm is more likely to spill into the originating firm itself as a form of proprietary knowledge (i.e., patents) after being extended, enhanced, and validated by support from external researchers. With a focus on the recent development of AI, we find several interesting conditions that facilitate the mechanism that knowledge spilled over from an originating firm spills into the firm as proprietary knowledge. In addition, we compare these conditions with those of external actors. Beyond the previous understanding that firms' publishing is an instrument for managing human resources, interacting with academia, and/or signaling R&D capabilities, this study provides novel evidence on the consequences of publication by firms from the viewpoint of knowledge spillovers. Moreover, the comparison of the patents gained by the originating firm and external actors provides several important managerial and theoretical implications.

6.1 Firm-level implications

We find that the size of the KSP is more positively related to the originating firm's creation of proprietary knowledge connected to the KSP than to the knowledge creation of external actors. In addition, the proportion of industry papers in the KSP has a positive impact on the gaining of proprietary knowledge by both external actors and the originating firm, with no significant difference in this effect between the two groups. Lastly, we find that the similarity between a focal paper and its KSP has a positive effect on the proprietary knowledge gained by both groups, whereas the size of the effect is significantly higher for the focal firm than for external actors. These results imply that research publication can be a strategic tool when research outcomes are selectively published by firms, considering the size and type of the audience expected to be interested in a research topic.

Specifically, the results imply that a firm is more likely to create proprietary knowledge linked to the KSP when the published knowledge is enhanced by more researchers and when contributing researchers have highly similar research interests to those of the focal firm. These findings indicate that the focal firm's efforts to predict the external actors expected to contribute to the KSP formulation would be useful for deciding which knowledge to selectively reveal. Given that the focal firm is in the best position to understand the potential pool of researchers who have the capability to extend the focal firm's knowledge, we can expect such prediction efforts to be reasonably guided by the focal firm's existing knowledge background. The experiences of relevant experts, as well as data mining and analytical skills based on bibliographic and patent databases, would be helpful for managing these prediction efforts systematically. After deciding what to publish, the originating firm can also promote its research to targeted external researchers who can extend the focal firm's work. Targeted promotion can ultimately facilitate an increase in the size of the KSP and the similarity between the revealed idea and KSP. Furthermore, by observing the evolution of the KSP, the focal firm can refine its search behavior to take advantage of its revealing strategy (Dosi, 1988).

By contrast, when the proportion of industry papers in the KSP increases, this is not always interpreted as a good sign for the originating firm because external actors' creation of proprietary knowledge increases similarly. Therefore, for an originating firm, a high proportion of industry papers in the KSP can be regarded as beneficial only when it has competitive advantage in terms of the key assets that complement the revealed knowledge. If an originating firm lacks the crucial assets needed to achieve commercialization, publications that form the KSP consisting of many firm papers may benefit external actors in the long run rather than the originating firm. If both the originating firm and external actors have their own complementary resources, it would be helpful for the originating firm to try to gain relevant patents as soon as possible because the number of potential competitors in that emerging area may increase rapidly over time.

To summarize, before publishing on an emerging science-related technology, a firm should monitor the number of external researchers to advance knowledge close to its research interests as well as whether such researchers are affiliated with firms or universities. Moreover, a firm should check whether it possesses any of the key assets needed to capture the value from the KSP of the published idea.

6.2 Contributions to the literature

This study offers a balanced view of the literature on research publications by firms. Studies in this strand have repeatedly mentioned the risk of revealing research outcomes to the public because of the potential damage from spillovers. On the contrary, firms that reveal knowledge can enjoy several positive consequences such as learning from external actors that respond to the knowledge revealed by the firm (Agarwal *et al.*, 2007). Focusing on the creation of patents by the focal firm and external actors linked to the KSP, we propose a dynamic as well as a balanced explanation of the consequences of firms' publishing.

This study contributes to the open innovation literature, particularly advancing the concept of selective revealing. Firms' publishing is a type of selective revealing strategy that discloses solution-related knowledge to create a new knowledge path. Case studies show that revealing a solution can serve as a strategy to attract others' support and thereby create new paths in nascent areas (Phillips *et al.*, 2000; Garud *et al.*, 2002; Dodgson *et al.*, 2007). By encouraging external actors to use and advance the revealed knowledge, the revealing firm can thus increase the possibility of shaping its environment and shifting toward a trajectory it favors. We extend this previous discussion by suggesting that solution revealing in an emerging science-related technology can be a possible route of learning from the ideas generated by external researchers, beyond simply shaping their participation in the relevant area.

The findings of this study can be linked to the evolutionary perspective of technology development. Some scholars have argued that the indirect benefits of selective revealing can outweigh the cost of the focal firm (e.g., Alexy *et al.*, 2013) because of the path-dependent and cumulative nature of knowledge (Nelson and Winter, 1982). Our empirical findings extend previous theoretical speculation by showing the specific conditions under which spillovers from the focal firm's publication can form a more idiosyncratic pool of knowledge to the focal firm, where the focal firm is more likely to take advantage. Knowledge in the KSP is likely to be linked to more patent applications by the focal firm when the KSP is larger than or similar to the focal paper.

Our findings contradict the conventional approaches used in the knowledge spillover literature (e.g., Jaffe, 1986; Jaffe *et al.*, 1993), which assume that the availability of knowledge spillovers is homogeneous for firms within a particular boundary such as region and sector. We show that the spillovers created by selective revealing can form a pool of focal firm-specific knowledge, which can be better exploited by the focal firm depending on how the revealed knowledge evolves over time. Therefore, if a firm can better predict the scope and characteristics of potential external actors who will reciprocate, selective revealing can be used as a more effective strategic tool to improve its technological position when science-related technology emerges.

6.3 Limitations and future research

Although these findings suggest several implications for both academia and practice, this study has several limitations that must be addressed in future work. First, it focuses on gaining proprietary knowledge through knowledge spillovers to understand the consequences of corporate publishing. Other dimensions of consequences such as attracting talented researchers and reaping monetary benefits are also important matters of debate in the long run. Future empirical research is thus needed to explore the other aspects of consequences depending on the major goals of the publishing firm.

The meaning of a patent citing a prior research publication can be interpreted differently from that of our research. We see prior knowledge as a source of inspiration for future patents. However, prior art can deter future patent application by challenging patentability (Della Malva and Hussinger, 2012). Therefore, although we regard external actors' patent creation linked to a KSP as learning something from the focal firm's publishing, the focal firm's publishing could have

deterred the patentability of external actors' inventions.¹¹ In a similar vein, previous studies have also mentioned that firms publish for defensive purposes, which blocks other actors' patenting by building prior art (Bar-Gill and Parchomovsky, 2003; Johnson, 2014). Given that firms can publish research for several reasons simultaneously, these different interpretations can coexist (i.e., they are not mutually exclusive). Therefore, it would be fruitful for future research on the consequences of corporate publishing to address the other meaning of patent applications that cite prior research, focusing more on the defensive purpose of corporate publishing.

Furthermore, although we control for the features related to the focal paper and focal firm itself, this study focuses on KSP-related features. As the focal firm does not directly control the KSP, our findings are less actionable. Nevertheless, we argue that firms' efforts to predict the KSP formulation can be a useful and reasonable approach for the strategic revealing of knowledge because the focal firm is in a relatively advantageous position to understand who has the capability to advance the focal firm's knowledge and how the KSP will be shaped. In addition, we suggest viable strategies after publishing, such as adjusting search behavior by observing the evolution of the KSP or running a targeted promotion of published topics to external researchers equipped with the relevant skills and knowledge. Future studies could complement our approach by directly investigating controllable attributes to derive more actionable suggestions.

Another potential limitation is the generalizability of our findings. The recent development of AI provides an appropriate context within which to observe active research publications by incumbent firms as well as explore the citation linkages between research papers and patents. Moreover, firms that publish their research usually have complementary assets, particularly data, when doing so. It does not seem likely that other fields that do not satisfy those conditions would also show similar findings to those of this study. However, at the same time, one can say that generalizing the findings may be misleading. Instead, it is more reasonable to discuss the details of the boundary conditions so that one can anticipate similar outcomes to those herein. Such an argument relies on the evolutionary view on innovation, which emphasizes a significant variation in the evolutionary patterns among industries (Pavitt, 1984). Such evolutionary patterns are related to technologies, firms, universities, and other interacting actors as well as the cultures and norms of the particular industry. From this viewpoint, our findings can be interpreted as outcomes obtained under particular constraints characterizing the field of study (i.e., AI), such as the unique publication culture, engagement in research publication by large incumbent firms, and the blurred boundary between science and technology. Therefore, future efforts are needed to understand different (or similar) consequences from firms' publishing in other emerging areas where firms do publish.

Although we control for variables that could be related to the quality of the published firm papers, such as the authors' overall performance and size of the affiliated organization, the innate quality of the paper itself is not directly reflected in the model. This may result in omitted variable bias because high-quality papers are likely to create a larger KSP, making it more likely that the KSP paper is cited not necessarily because inventors learn from the paper but because they must cite it. One way to check this possibility is to investigate the share of examiner-added citations, which represent citations added by examiners rather than inventors. By definition, examiner-added citations are likely to be less relevant to actual learning than citations added by inventors. In our data, fewer than 1% of the citations are examiner-added citations, which is in line with the literature showing that the proportion of examiner-added citations is negligible for non-patent citations (e.g., Lemley and Sampat, 2012; Ahmadpoor and Jones, 2017; Bikard and Marx, 2020). This mitigates concerns about omitted variable bias because the citation linkages used in our analysis are more likely to represent an authentic flow of knowledge.

Lastly, this study follows Arora *et al.* (2018) to define a firm paper (i.e., at least one of the authors is firm-affiliated). The engagement of a firm-affiliated author is likely to imply that the

¹¹ Cited documents showing features that question the novelty of the claimed invention when taken alone are called X-type citations. In our data, about 19% of citations from patents to the KSP are X-type citations. Given the existence of defensive publishing (although various motivations coexist), the share of X-type citations can be higher than this if one examines the direct citation linkages between focal firm publication and subsequent patent applications (e.g., Della Malva and Hussinger, 2012).

corresponding paper addresses topics related to the firm's interests and is developed by exploiting the firm's infrastructure. Nevertheless, such a definition is a simplified way of defining a firm paper, given that the actual contribution of firm-affiliated authors can vary across papers. Depending on the level of engagement by a firm, spillovers can occur in a different direction, thereby potentially bringing about different consequences when creating proprietary knowledge. Future studies could additionally reflect on such aspects by considering the degree of engagement of the firm-affiliated authors in each paper.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C2005026). The authors are grateful to Ben Martin, Paul Nightingale, and Ed Steinmueller at the Science Policy Research Unit, University of Sussex for their helpful comments on the early draft of this manuscript. This study is a revised version of the first author's doctoral thesis chapter supervised by the second author at Yonsei University.

References

- Abernathy, W. J. and J. M. Utterback (1978), 'Patterns of industrial innovation,' *Technology Review*, 80(7), 40–47.
- Agarwal, H., D. Audretsch and M. B. Sarkar (2007), 'The process of creative construction: knowledge spillovers, entrepreneurship, and economic growth,' *Strategic Entrepreneurship Journal*, 1(3–4), 263–286.
- Agarwal, R., D. Audretsch and M. B. Sarkar (2010), 'Knowledge spillovers and strategic entrepreneurship,' *Strategic Entrepreneurship Journal*, 4(4), 271–283.
- Ahmadpoor, M. and B. F. Jones (2017), 'The dual frontier: patented inventions and prior scientific advance,' *Science*, 357(6351), 583–587.
- Alexy, O., P. Criscuolo and A. Salter (2012), 'Managing unsolicited ideas for R&D,' *California Management Review*, 54(3), 116–139.
- Alexy, O., G. George and A. J. Salter (2013), 'Cui bono? The selective revealing of knowledge and its implications for innovative activity,' *Academy of Management Review*, 38(2), 270–291.
- Allen, T. J. (1984), *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*. MIT Press Books: Cambridge MA, p. 1.
- Anderson, P. and M. L. Tushman (1990), 'Technological discontinuities and dominant designs: a cyclical model of technological change,' *Administrative Science Quarterly*, 35(4), 604–633.
- Anthes, G. (2017), 'Artificial intelligence poised to ride a new wave,' *Communications of the ACM*, 60(7), 19–21.
- Arora, A., S. Belenzon and A. Pataconi (2018), 'The decline of science in corporate R&D,' *Strategic Management Journal*, 39(1), 3–32.
- Arrow, K. (1962), 'Economic welfare and the allocation of resources for invention,' in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity*. Princeton University Press: Princeton, NJ, pp. 609–626.
- Arundel, A. (2001), 'The relative effectiveness of patents and secrecy for appropriation,' *Research Policy*, 30(4), 611–624.
- Audretsch, D. B. and M. P. Feldman (1996), 'R&D spillovers and the geography of innovation and production,' *The American Economic Review*, 86(3), 630–640.
- Bar-Gill, O. and G. Parchomovsky (2003), 'The value of giving away secrets,' *Virginia Law Review*, 89(8), 1857–1895.
- Barney, J. (1991), 'Firm resources and sustained competitive advantage,' *Journal of Management*, 17(1), 99–120.
- Baruffaldi, S. and F. Poege (2020), 'A firm scientific community: industry participation and knowledge diffusion,' *Max Planck Institute for Innovation & Competition Research Paper* (20–10).
- Basberg, B. L. (1987), 'Patents and the measurement of technological change: a survey of the literature,' *Research Policy*, 16(2–4), 131–141.
- Bikard, M. and M. Marx (2020), 'Bridging academia and industry: how geographic hubs connect university science and corporate technology,' *Management Science*, 66(8), 3425–3443.
- Blei, D. M. and J. D. Lafferty (2006), 'Dynamic topic models,' *Proceedings of the 23rd International Conference on Machine Learning*, ACM, pp. 113–120.
- Blei, D. M. and J. D. Lafferty (2007), 'A correlated topic model of science,' *Annals of Applied Statistics*, 1(2), 634–642.
- Boden, M. A. (2016), *AI: Its Nature and Future*. Oxford University Press.

- Bridgstock, M. (1991), 'The quality of single and multiple authored papers: an unresolved problem,' *Scientometrics*, 21(1), 37–48.
- Chang, J., J. Boyd-Graber, C. Wang, S. Gerrish and D. Blei (2009), 'Reading tea leaves: how humans interpret topic models,' *Proceedings of Neural Information Processing Systems*. Vancouver, B.C., Canada.
- Cohen, W. M., R. R. Nelson and J. P. Walsh (2000), 'Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent (or not) (No. w7552),' *National Bureau of Economic Research*.
- Cohen, W. M. and D. A. Levinthal (1990), 'Absorptive capacity: A new perspective on learning and innovation,' *Administrative Science Quarterly*, 35(1), 128–152.
- Cook, R. J. and J. Lawless (2007), *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media: New York.
- Cox, D. R. (1972), 'Regression models and life-tables,' *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.
- Dahlander, L. and D. M. Gann (2010), 'How open is innovation?' *Research Policy*, 39(6), 699–709.
- Dasgupta, P. and P. A. David (1994), 'Toward a new economics of science,' *Research Policy*, 23(5), 487–521.
- Della Malva, A. and K. Hussinger (2012), 'Corporate science in the patent system: an analysis of the semiconductor technology,' *Journal of Economic Behavior & Organization*, 84(1), 118–135.
- Demsetz, H. (1967), 'Towards a theory of property rights,' *American Economic Review*, 57, 347–359.
- Denning, P. J. (2005), 'Is computer science science?' *Communications of the ACM*, 48(4), 27–31.
- Didegah, F. and M. Thelwall (2013), 'Which factors help authors produce the highest impact research? Collaboration, journal and document properties,' *Journal of Informetrics*, 7(4), 861–873.
- Dodgson, M., D. M. Gann and A. Salter (2007), "'In case of fire, please use the elevator": simulation technology and organization in fire engineering,' *Organization Science*, 18(5), 849–864.
- Dosi, G. (1982), 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change,' *Research Policy*, 11(3), 147–162.
- Dosi, G. (1988), 'Sources, procedures, and microeconomic effects of innovation,' *Journal of Economic Literature* 26(3), 1120–1171.
- Dyer, J. H. and H. Singh (1998), 'The relational view: cooperative strategy and sources of interorganizational competitive advantage,' *Academy of Management Review*, 23(4), 660–679.
- Etzkowitz, H. (2003), 'Research groups as 'quasi-firms': the invention of the entrepreneurial university,' *Research Policy*, 32(1), 109–121.
- Fleming, L. (2001), 'Recombinant uncertainty in technological search,' *Management Science*, 47(1), 117–132.
- Foss-Solbrekk, K. (2021), 'Three routes to protecting AI systems and their algorithms under IP law: the good, the bad and the ugly,' *Journal of Intellectual Property Law & Practice*, 16(3), 247–258.
- Franceschet, M. and A. Costantini (2010), 'The effect of scholar collaboration on impact and quality of academic papers,' *Journal of Informetrics*, 4(4), 540–553.
- Freeman, C. (1997), *The Economics of Industrial Innovation*, 3rd edn. Routledge: London.
- Freyne, J., L. Coyle, B. Smyth and P. Cunningham (2010), 'Relative status of journal and conference publications in computer science,' *Communications of the ACM*, 53(11), 124–132.
- Garfield, E. (1979), *Citation Indexing. Its Theory and Application in Science, Technology and Humanities*. Wiley: New York.
- Garud, R., S. Jain and A. Kumaraswamy (2002), 'Institutional entrepreneurship in the sponsorship of common technological standards: the case of Sun Microsystems and Java,' *Academy of Management Journal*, 45(1), 196–214.
- Garud, R. and M. Rappa (1994), 'A socio-cognitive model of technology evolution,' *Organization Science*, 5(3), 344–362.
- Gazni, A. and F. Didegah (2011), 'Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications,' *Scientometrics*, 87(2), 251–265.
- Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994), *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. Sage: London.
- Granstrand, O. (1999), *The Economics and Management of Intellectual Property: Towards Intellectual Capitalism*. Edward Elgar: Cheltenham.
- Granstrand, O. (2018), *Evolving Properties of Intellectual Capitalism: Patents and Innovations for Growth and Welfare*. Edward Elgar: Cheltenham.
- Grant, R. M. (1996), 'Prospering in dynamically-competitive environments: organizational capability as knowledge integration,' *Organization Science*, 7(4), 375–387.
- Grassano, N., R. Camerani and D. Rotolo (2019), 'Paper to be presented at DRUID19 Copenhagen Business School,' Copenhagen, Denmark June 19–21, 2019.
- Griliches, Z. (1991), *The Search for R&D Spillovers* (No. W3768). National Bureau of Economic Research.
- Hall, D., D. Jurafsky and C. D. Manning (2008), 'Studying the histories of ideas using topic models,' *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii.

- Hartmann, P. and J. Henkel (2020), 'The rise of corporate science in AI: data as a strategic resource,' *Academy of Management Discoveries*, 6(3), 359–381.
- Hayter, C. S. and A. N. Link (2018), 'Why do knowledge-intensive entrepreneurial firms publish their innovative ideas?' *Academy of Management Perspectives*, 32(1), 141–155.
- Heckman, J. J. and G. J. Borjas (1980), 'Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence,' *Economica*, 47(187), 247–283.
- Henkel, J. (2006), 'Selective revealing in open innovation processes: the case of embedded Linux,' *Research Policy*, 35(7), 953–969.
- Henkel, J., S. Schöberl and O. Alexy (2014), 'The emergence of openness: how and why firms adopt selective revealing in open innovation,' *Research Policy*, 43(5), 879–890.
- Hicks, D. (1995), 'Published papers, tacit competencies and corporate management of the public/private character of knowledge,' *Industrial and Corporate Change*, 4(2), 401–424.
- Hinton, G. E., S. Osindero and Y. W. Teh (2006), 'A fast learning algorithm for deep belief nets,' *Neural Computation*, 18(7), 1527–1554.
- Jaffe, A. B. (1986), 'Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value,' *American Economic Review*, 76, 984–1001.
- Jaffe, A. B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations,' *The Quarterly Journal of Economics*, 108(3), 577–598.
- Jee, S. J., M. Kwon, J. M. Ha and S. Y. Sohn (2019), 'Exploring the forward citation patterns of patents based on the evolution of technology fields,' *Journal of Informetrics*, 13(4), 100985.
- Jee, S. J. and S. Y. Sohn (2022), 'Firms' influence on the evolution of published knowledge when a science-related technology emerges: the case of artificial intelligence,' *Journal of Evolutionary Economics*, 1–39.
- Johnson, J. P. (2014), 'Defensive publishing by a leading firm,' *Information Economics and Policy*, 28, 15–27.
- Kalbfleisch, J. D. and R. L. Prentice (1980), *The Statistical Analysis of Failure Time Data*. Wiley: New York.
- Kale, P. and H. Singh (2009), 'Managing strategic alliances: what do we know now, and where do we go from here?' *The Academy of Management Perspectives* 23(3), 45–62.
- Kaplan, S. and K. Vakili (2015), 'The double-edged sword of recombination in breakthrough innovation,' *Strategic Management Journal*, 36(10), 1435–1457.
- Katz, J. S. and B. R. Martin (1997), 'What is research collaboration?' *Research Policy*, 26(1), 1–18.
- Kogut, B. and U. Zander (1992), 'Knowledge of the firm, combinative capabilities, and the replication of technology,' *Organization Science*, 3(3), 383–397.
- Kotha, R., G. George and K. Srikanth (2013), 'Bridging the mutual knowledge gap: coordination and the commercialization of university science,' *Academy of Management Journal*, 56(2), 498–524.
- Lane, P. J. and M. Lubatkin (1998), 'Relative absorptive capacity and interorganizational learning,' *Strategic Management Journal*, 19(5), 461–477.
- Lemley, M. A. and B. Sampat (2012), 'Examiner characteristics and patent office outcomes,' *Review of Economics and Statistics*, 94(3), 817–827.
- Makri, M., M. A. Hitt and P. J. Lane (2010), 'Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions,' *Strategic Management Journal*, 31(6), 602–628.
- Marco, A. C. (2007), 'The dynamics of patent citations,' *Economics Letters*, 94(2), 290–296.
- Merton, R. K. (1968), 'The Matthew effect in science: the reward and communication systems of science are considered,' *Science*, 159(3810), 56–63.
- Meyer-Krahmer, F. and U. Schmoch (1998), 'Science-based technologies: university–industry interactions in four fields,' *Research Policy*, 27(8), 835–851.
- Mowery, D. C., J. E. Oxley and B. S. Silverman (1996), 'Strategic alliances and interfirm knowledge transfer,' *Strategic Management Journal*, 17(S2), 77–91.
- Murray, F. (2002), 'Innovation as co-evolution of scientific and technological networks: exploring tissue engineering,' *Research Policy*, 31(8–9), 1389–1403.
- Narin, F., K. S. Hamilton and D. Olivastro (1997), 'The increasing linkage between US technology and public science,' *Research Policy*, 26(3), 317–330.
- Nelson, R. and S. G. Winter (1982), *An Evolutionary Theory of Economic Change*. Harvard University Press: Cambridge.
- Nerkar, A. and S. Paruchuri (2005), 'Evolution of R&D capabilities: the role of knowledge networks within a firm,' *Management Science*, 51(5), 771–785.
- Norton, S., (2016), 'CIO explainer: what is artificial intelligence?' wsj.com, July 18. <http://blogs.wsj.com/cio/2016/07/18/cio-explainer-what-is-artificial-intelligence/>.
- Nuvolari, A. (2004), 'Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine,' *Cambridge Journal of Economics*, 28(3), 347–363.

- Owen-Smith, J. and W. W. Powell (2004), 'Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community,' *Organization Science*, 15(1), 5–21.
- Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory,' *Research Policy*, 13(6), 343–373.
- Pavitt, K. (1987), 'The objectives of technology policy,' *Science & Public Policy*, 14(4), 182–188.
- Phillips, N., T. B. Lawrence and C. Hardy (2000), 'Inter-organizational collaboration and the dynamics of institutional fields,' *Journal of Management Studies*, 37(1), 22–43.
- Podolny, J. M. and T. E. Stuart (1995), 'A role-based ecology of technological change,' *American Journal of Sociology*, 100(5), 1224–1260.
- Polidoro, F. and M. Theeke (2012), 'Getting competition down to a science: the effects of technological competition on firms' scientific publications,' *Organization Science*, 23(4), 1135–1153.
- Polidoro, F. and P. K. Toh (2011), 'Letting rivals come close or warding them off? The effects of substitution threat on imitation deterrence,' *Academy of Management Journal*, 54(2), 369–392.
- Romer, P. (1990), 'Endogenous technological change,' *Journal of Political Economy*, 98(5, Part 2), S71–108.
- Rosenberg, N. (1990), 'Why do firms do basic research (with their own money)?' *Research Policy*, 19(2), 165–174.
- Rosenberg, N. and R. R. Nelson (1994), 'American universities and technical advance in industry,' *Research Policy*, 23(3), 323–348.
- Simeth, M. and M. Cincera (2016), 'Corporate science, innovation, and firm value,' *Management Science*, 62(7), 1970–1981.
- Sinha, A., Z. Shen, Y. Song, H. Ma, D. Eide, B. J. Hsu and K. Wang (2015), 'An overview of Microsoft Academic Service (MAS) and applications,' *Proceedings of the 24th International Conference on World Wide Web*. Florence: Italy, pp. 243–246.
- Smart, J. and A. Bayer (1986), 'Author collaboration and impact: a note on citation rates of single and multiple authored articles,' *Scientometrics*, 10(5–6), 297–305.
- Somaya, D. (2012), 'Patent strategy and management: an integrative review and research agenda,' *Journal of Management*, 38(4), 1084–1114.
- Taddy, M. (2018), 'The technological elements of artificial intelligence (No. w24301),' *National Bureau of Economic Research*.
- Teece, D. J. (1986), 'Profiting from technological innovation: implications for integration, collaboration, licensing and public policy,' *Research Policy*, 15(6), 285–305.
- Teece, D. J. (2018), 'Profiting from innovation in the digital economy: enabling technologies, standards, and licensing models in the wireless world,' *Research Policy*, 47(8), 1367–1387.
- USPTO IP Data Highlights. (2020), 'Inventing AI: tracing the diffusion of artificial intelligence with US patents,' <https://www.uspto.gov/sites/default/files/documents/OCE-DH-AI.pdf>.
- Van Looy, B., T. Magerman and K. Debackere (2007), 'Developing technology in the vicinity of science: an examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology,' *Scientometrics*, 70(2), 441–458.
- Vardi, M. Y. (2009), 'Conferences vs. journals in computing research,' *Communications of the ACM*, 52(5), 5.
- Von Hippel, E. (1988), *The Sources of Innovation*. Oxford University Press: New York.
- Von Krogh, G., S. Haefliger, S. Spaeth and M. W. Wallin (2012), 'Carrots and rainbows: motivation and social practice in open source software development,' *MIS Quarterly*, 36(2), 649–676.
- Wernerfelt, B. (1984), 'A resource-based view of the firm,' *Strategic Management Journal*, 5(2), 171–180.
- WIPO Technology Trends. (2019), 'Artificial intelligence,' <https://www.wipo.int/publications/en/details.jsp?id=4386>.
- Yang, H., C. Phelps and H. K. Steensma (2010), 'Learning from what others have learned from you: the effects of knowledge spillovers on originating firms,' *Academy of Management Journal*, 53(2), 371–389.
- Zahra, S. A. and G. George (2002), 'Absorptive capacity: A review, reconceptualization, and extension,' *Academy of Management Review*, 27(2), 185–203.
- Zucker, L. (1987), 'Institutional theories of organizations,' *Annual Review of Sociology*, 13(1), 443–464.

Appendix A

Table A1. Major conferences and journals

Conferences	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
	Neural Information Processing Systems (NIPS)
	International Conference on Machine Learning (ICML)
	IEEE International Conference on Computer Vision (ICCV)
	International Conference on Knowledge Discovery and Data Mining (SIGKDD)
	Meeting of the Association for Computational Linguistics (ACL)
	ACM SIGMOD International Conference on Management of Data
	Conference on Empirical Methods in Natural Language Processing (EMNLP)
	AAAI Conference on Artificial Intelligence (AAAI)
Journals	IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
	<i>IEEE Transactions on Evolutionary Computation</i>
	<i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i>
	<i>IEEE Transactions on Fuzzy Systems</i>
	<i>IEEE Computational Intelligence Magazine</i>
	<i>International Journal of Neural Systems</i>
	<i>Information Fusion</i>
	<i>Automatica</i>
	<i>Neural Networks</i>
	<i>Journal of Machine Learning Research</i>
	<i>Information Sciences</i>

Source: [Jee and Sohn \(2022\)](#).

Appendix B

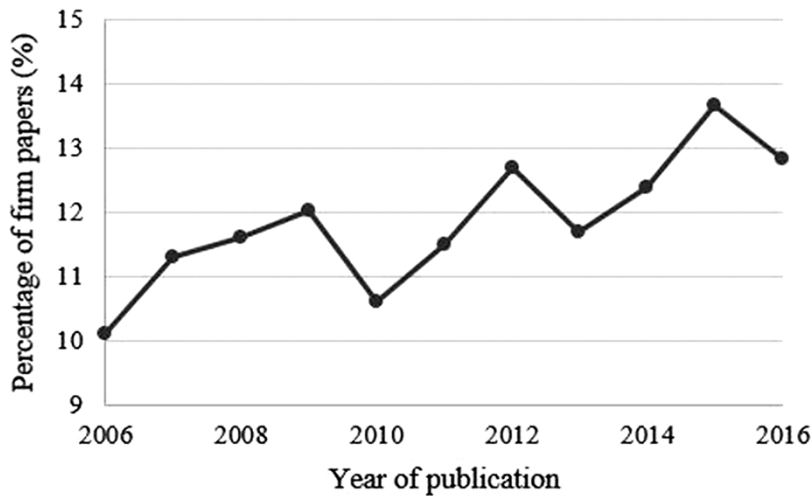


Figure B1. Share of firm papers in all the papers from the conferences and journals listed in [Table A1](#) Note: Firm papers in this figure cover the publications by firms we address, as described in [Section 5.1](#). The figure shows a gradually increasing trend in the proportion of firm papers in this field despite a significant increase in the total number of publications in this field as AI emerges ([WIPO Technology Trends, 2019](#)).

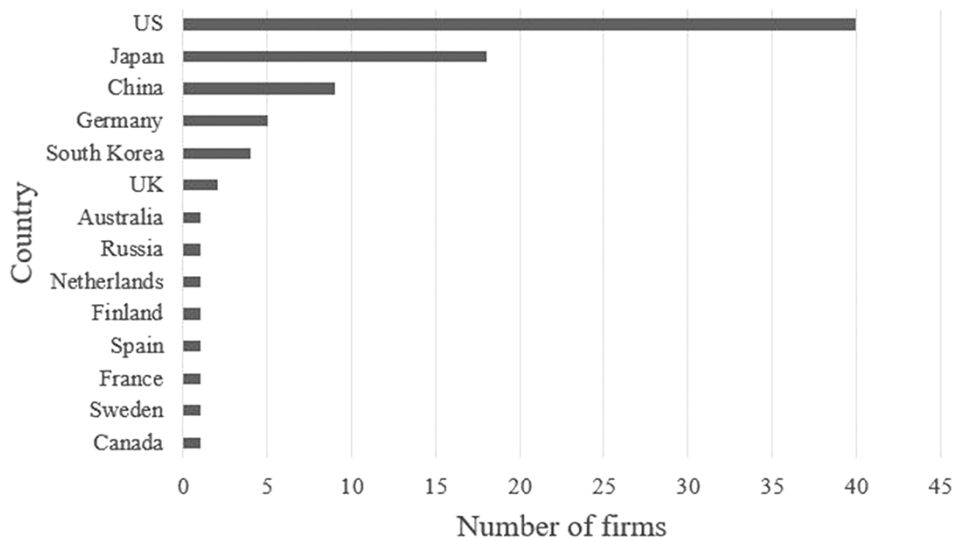


Figure B2. Number of firms publishing papers in the selected conferences and journals by country

Table B1. List of major firms publishing on AI

Firm name	Number of papers published in the selected conferences and journals (2006–2016)
Microsoft	1811
Google	838
IBM	809
Yahoo	452
Siemens	240
Toyota	224
Intel	203
Adobe	166
Honda	164
NEC	156
Facebook	126
NTT Data	112
Samsung	109
Xerox	106
Mitsubishi	93

Yahoo! is excluded from our formal analysis, as its firm size information is not available from Compustat, presumably because of its severely low performance since the mid-2000s.

Table B2. List of major organizations (excluding firms) publishing on AI

Organization name	Number of papers published in the selected conferences and journals (2006–2016)
Carnegie Mellon University	1986
Massachusetts Institute of Technology	1221
Stanford University	1025
University of California Berkeley	987
Chinese Academy of Sciences	904
University of Illinois at Urbana Champaign	850
National University of Singapore	777

(continued)

Table B2. (Continued)

Organization name	Number of papers published in the selected conferences and journals (2006–2016)
Tsinghua University	773
Max Plank Society	709
French Institute for Research in Computer Science and Automation	709
University of Tokyo	699
University of Southern California	689
ETH Zurich	682
University of Washington	676
University of Texas at Austin	645

These numbers are obtained from the Microsoft Academic Graph database, which provides affiliation-specific IDs with high accuracy (Sinha *et al.*, 2015).