Journal of the Knowledge Economy (2024) 15:20547–20577
https://doi.org/10.1007/s13132-024-01974-1

The Over-Concentration of Innovation and Firm-Specific Knowledge in the Artificial Intelligence Industry

Pedro Jácome de Moura Jr.[1] · **Carlos Denner dos Santos Junior**[2] · **Carlo Gabriel Porto-Bellini**[1] · **José Jorge Lima Dias Junior**[1]

Received: 6 February 2023 / Accepted: 4 April 2024 / Published online: 16 April 2024

Abstract
The development of the artificial intelligence (AI) landscape has been impressive in virtually all economic sectors in recent years. Our study discusses the over-concentration of AI knowledge (OCAIK) as the origin of dominance over the global AI industry by a small number of companies and universities that deploy the needed resources to develop and use cutting edge, inimitable AI knowledge. Business agents appropriate AI-related scholarly research and absorb research findings that grant them increasingly inimitable competitive advantages over new entrants. Our study verifies the occurrence of OCAIK by processing thousands of papers presented in AI conferences from 2013 to 2022. To analyze our hypotheses, we used classification techniques and inferential statistics. We found a significant difference between clusters of companies that we called ordinary investors and outlier investors. We also observed the influence of universities in the correlation between OCAIK and investments made in both research and development (R&D) and capital goods. Our findings indicate a strong collaboration between AI leading companies and universities in generating firm-specific AI knowledge. We additionally offer novel insights on the resource-based view (RBV) and the knowledge-based view (KBV) research traditions, in that business competition may reach a point of no return if only incremental innovation is devised instead of radical innovation to break the chains of knowledge accumulation and technological implementation by a strict number of agents.

Keywords Artificial intelligence · Competitive advantage · Research and development · Knowledge-based view · Oligopolies

This article is part of the Topical Collection on *AI in the Knowledge Economy and Society: Implications for Theory, Policy and Practice*

Extended author information available on the last page of the article

Introduction

The world has approximately 214 million companies in various fields of industry, agriculture, trade, and services (Statista, 2022). They compete by means of differentiation to gain relative advantage over competitors and, consequently, market share. One of the main strategic actions to obtain competitive advantage has been to invest in research and development (R&D) (Boiko, 2021; Menke, 1997; Nayak et al., 2022). About a trillion dollars is invested annually in R&D by the 2500 companies that lead R&D in the world, corresponding to 90% of the world's business-funded R&D (Grassano et al., 2022). Knowledge has been long considered resourceful to leverage the needed differentiation and, ultimately, competitiveness (Barney, 1991; Grant, 1996). In particular, firm-specific knowledge is better for business advantage than publicly available knowledge, since the proprietary nature of firm-specific knowledge makes it more difficult to absorb, imitate, or replace (Wang et al., 2016). Technology thus emerges as key to acquire and exploit firm-specific knowledge (Nayak et al., 2022), particularly the technology that allows for big data analytics (Dahiya et al., 2021) and artificial intelligence (AI) (Fredström et al., 2021).

Towards competitive advantage, companies may want to establish partnership with universities as universities have the necessary background to conduct research in emerging, knowledge-based fields such as AI (Shao et al., 2020). Moreover, given the focus of universities on teaching, basic research, and community outreach, i.e., they assume a more non-commercial mission, universities do not represent serious threats to businesses (Choi & Contractor, 2019). Also, to exploit business advantages, universities face greater difficulties to transform resources and skills into real capabilities or core competences (Mahdi et al., 2019). The latter may be due to the fact that while business companies focus on the integration of resources (e.g., research findings) and the exploitation of advantages acquired by the combination of resources, universities in their turn focus on the production of factual knowledge (to promote an open debate within the academic community as well as to advance the frontiers of scholarly knowledge). Moreover, universities are not necessarily interested in the application of integrated resources. To illustrate this point, take Microsoft's *Optics for the Cloud* project, which aims to revolutionize data storage in partnership with multiple US and UK universities (Parmigiani et al., 2021), where each university carries out specific studies and whose results are integrated with results from other universities under Microsoft's discretion and power. This is an example that universities work in the periphery of business initiatives. The peripheral role in the business landscape is a consequence of companies seeking profit and competitive advantage, whereas universities focus on research rather than on marketing a product (D'Este & Perkmann, 2010). Another reason for the peripheral role of universities is that even when universities focus their entrepreneurial activities on technology licensing through spin-off firms, the spin-offs are often bought soon after by large companies as a way to gain access to cutting-edge knowledge (Ferreira et al., 2018).

AI, in particular, has received attention from governments and scholars regarding ethical issues, regulation, and governance (e.g., Kerr et al., 2020; Roberts et al., 2021). An example of governmental concern regarding AI ethics, regulation, and governance is the *AI Watch Report* launched by the European Commission in December 2018 to monitor the impact of AI as a source of potential social implications (Benetta et al., 2021). On the academic side, a related initiative is the *AI Governance: A Research Agenda Report*, published by the Future of Humanity Institute (FHI), University of Oxford, which highlights risks of a potential oligopolistic global market structure motivated by interests on AI (Dafoe, 2018). Such an oligopolistic structure manifests when only a few companies can invest what is needed regarding capital and R&D to keep pace with constant technical iterations (Ding & Dafoe, 2021). Such issues lead us to assume that companies' accumulated R&D investment capacity, when reinforced by capabilities generated through partnerships with universities, determines the level of firm-specific AI knowledge and competitive performance. The relationship between investments and competitive performance is in fact well documented in the information technology (IT) literature. For example, IT as a knowledge-intensive industry uses its capabilities on combining IT resources to create competitive advantages (Bharadwaj, 2000). Since resources cannot offer competitive advantages by themselves (Cohen & Olsen, 2013), knowledge, skills, and partnerships are examples of complementary capabilities that explain a firm's superior performance (Fink, 2011) and could possibly explain the current advances in AI. In fact, in the light of the knowledge-based view (KBV), knowledge, skills, and partnerships are needed for firm-specific knowledge integration into organizational capability (Grant, 1996), which makes KBV a promising theory for analyzing the *over-concentration of AI knowledge* (hereafter referred to as OCAIK).

While studies on AI investments have focused mostly on regional statistics (e.g., Jeon et al., 2024; Benetta et al., 2021; and Dafoe, 2018), the role of large companies in shaping the AI landscape has been anecdotal (e.g., Newman, 2017; Webb, 2019), leading us to elaborate the following research questions:

RQ1 To what extent do companies that invest heavily in firm-specific AI knowledge differentiate from those that do not?
RQ2 How do universities collaborate with companies to reinforce OCAIK?

By answering these questions, we expect to contribute in two ways for scholarly and applied knowledge. First, by discussing OCAIK, the study sheds light on the need to revisit the role of independently or publicly funded research institutions, so that those institutions become cognizant of OCAIK and champion policies for the regulation and democratization of access to AI resources. And second, the study adds to the KBV literature regarding how much advantage is needed for sustainable, competitive advantage.

We developed a parsimonious model to answer RQ1 and RQ2. Specifically, we test the positive relationship between (1) R&D investments by companies and the formation of a group of AI leading companies (LC); (2) R&D expenditure by universities and the formation of a group of AI leading universities (LU); (3) the partnership between LC and LU, and the production of LC firm-specific AI knowledge; and (4) the partnership between LC and universities in general, and the production of LC firm-specific AI knowledge.

The article is organized as follows. First, we discuss the theoretical framing of competitive advantage, the KBV, and the role of AI developments in that framing, especially machine learning. Second, we present a methodological approach that uses secondary data on R&D investments and academic production (papers published in leading AI conferences) to answer the research questions. Third, we discuss the results along with implications for theory and practice, limitations, and future studies. And fourth, we present conclusions about OCAIK and its potential consequences.

Literature Review

Competitive Advantage and the Knowledge-Based View of the Firm (KBV)

Competitive advantage is "a multifaceted construct arising out of diverse contextual manifestations" mostly regarding an economic perspective and grounded on rivalry and leadership (Nayak et al., 2022, p. 977). The concept has just recently gained a socio-technical perspective, as governments and industry become increasingly aware of societal expectations (Marakova et al., 2021) and of the role of emerging technologies for the competitive advantage at firm level (Shao et al., 2020) and country level (Ding & Dafoe, 2021). The resource-based view of the firm (RBV) comes to explain a firm's survival through its capacity of achieving differentiation in a competitive environment (Barney, 1991). In the light of the RBV, competitive advantage is achieved through cost reduction (e.g., scale gains, exclusive sources of raw materials, and a cost-effective workforce) or because of specific attributes of the firm's products or processes (e.g., trademarks, patents, network service, or distribution channels) (Newbert, 2008). The RBV assumes that (1) firms are composed of heterogeneous resources (the resources are not evenly distributed among all players); (2) firms will have superior performance if their resources are valuable, rare, difficult to imitate, irreplaceable, or difficult to move; and (3) the source of competitive advantage exists inside the firm. In their search for differentiation, firms develop organizational capabilities, i.e., a group of activities based on the development, flow, and exchange of information carried out systematically and allowing the firm to take advantage of its resources to generate valuable outcomes (Degravel, 2011). As such, the RBV expects certain stability in a competitive sector to give firms time to change and adapt (Teece et al., 1997).

However, in rapidly changing and unpredictable situations, knowledge emerges as the main source of competitive advantage by promoting the transformation of organizational capabilities into dynamic capabilities, i.e., "the organizational and strategic routines by which firms achieve new resource configurations as markets emerge, collide, split, evolve, and die" (Eisenhardt & Martin, 2000, p. 1107). The knowledge-based view of the firm (KBV) therefore extends the RBV by assuming that the need for dynamic capabilities is inherent to organizations and that knowledge is a superior resource, one which combines/recombines resources, including those that are external to the organization (Bharadwaj, 2000; Grant, 1996). The assumption of knowledge as a relevant organizational asset implies a strategic decision and further actions to promote curiosity to learn, adaptation/improvement of what has been learned, and exploitation to obtain benefits—what is known as *absorptive capacity* (Henard & McFadyen, 2006). The KBV thus explains competitive advantage by means of absorptive capacity, i.e., knowledge integration into organizational and dynamic capabilities (Zahra et al., 2020).

The KBV has been used to analyze R&D investments and competitive advantage in areas such as corporate social responsibility and sustainability (Ullah & Arslan, 2022), cybersecurity (Mongeau & Hajdasinski, 2021), emerging economies and national competitiveness (Ge & Liu, 2022), and in innovation in the healthcare sector (Orlando et al., 2021). However, it is not any sort of knowledge that promotes competitive advantage. Firm-specific knowledge has more potential for differentiation among competitors than publicly accessible knowledge (Barney, 1991; Grant, 1996), since the proprietary nature of firm-specific knowledge makes it more difficult to absorb, imitate, or replace (Pereira & Bamel, 2021; Wang et al., 2016). Firm-specific knowledge is composed of (1) expertise and skills of a firm's employees, manifesting through "common language, relationships, or a sense of identification that exists among departments within a firm," and increasing on the basis of employee involvement with prior related projects (Mayer et al., 2012, p. 3); (2) the levels of financial and human-resource slack for R&D (Wang et al., 2016); and (3) external sources and partnerships (Cai et al., 2019). Slack resources are also desirable for firm-specific knowledge development, since they exceed the minimum necessary for organizational operation, and, as such, work as "a buffer resource" in times of change (Yiu et al., 2020, p. 1210).

Companies that invest heavily in firm-specific knowledge are now combining their traditional innovation process with AI technologies to gain competitive advantage and differentiation vis-à-vis those who invest less (Bai & Li, 2020), specifically by applying AI to the development of firm's knowledge with potential for innovation in business models, product and service new features, innovation structure, market performance, and innovation in supply chain management (Bahoo et al., 2023).

The University-Industry-Government Helix

Academic findings may generate unique knowledge to be transferred to society through scholarly publications, but also by means of patenting, spin-off startups, joint research, and consulting (D'Este & Perkmann, 2010). Such modes of knowledge transfer have been long encouraged through public policies and government incentives (Etzkowitz & Leydesdorff, 1995). In fact, social agents urge universities to participate in collaborative innovation processes and knowledge transfer for social and economic development (Johnston, 2019). The *triple helix* metaphor illustrates the interaction that is expected to occur between universities, industry, and government (Etzkowitz & Leydesdorff, 1995). It is interesting for companies to collaborate with universities as universities focus on basic research rather than on profit, and as they do not pose threats to competition (Miotti & Sachwald, 2003). Universities are also of value for governments as universities promote open innovation and have the capacity to create and transfer knowledge to both the business and the public sectors (Johnston, 2019). Universities have thus been considered ideal partners for integration with governments and with the private sector in a tri-lateral collaborative arrangement, particularly in the domain of AI (Mikhaylov et al., 2018).

The traditional role of universities in teaching and research has been changing as a consequence of the competitive pressure universities face in the global economy (Ferreira et al., 2018), the decreasing availability of public funding (Miotto et al., 2020), and the idea that governments could stimulate universities to become entrepreneurial agents (Etzkowitz & Leydesdorff, 1995). Universities would then incorporate entrepreneurship in their institutional routines to obtain economic returns from knowledge creation (Formica, 2002). However, there is a lack of understanding on how knowledge is transferred and how collaborative partnerships are established, i.e., how companies select universities to partner with, and vice versa, as well as what are the underlying reasons for the partnership (Johnston, 2019).

Artificial Intelligence

AI emerged in the late 1940s (Bruderer, 2016) in seminal propositions on how to determine whether a digital computer could think like a human (Turing, 1950) and on how a computer could play human games, “something of the nature of judgement, and considerable trial and error, rather than a strict, unalterable computing process” (Shannon, 1950, p. 256). Shortly after, those propositions began to appear in practice, as described by Samuel (1959) when he wrote a computer program “to behave in a way which, if done by human beings or animals, would be described as involving the process of learning,” that is, “[p]rogramming computers to learn from experience” (Samuel, 1959, p. 211). Still, in the early days, the *Eliza Effect* (Weizenbaum, 1966) introduced to the world further anthropomorphic characteristics of AI, particularly one that emulated human conversation.

Broadly, AI depends on the ability of machines to learn from experience, examples, and planned training, or, as more recently defined, “[a] computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E” (Mitchell, 1997, p. 2). The concept of AI encompasses (1) inputs (raw data, parameters), models (rules, heuristics, accuracy), and outcomes (predictions, decisions, trustworthiness) (Bui et al., 2020; Kaur et al., 2020); (2) technology and resources (hardware, software, data quality, security) (Challa et al., 2020; Hu et al., 2021); and (3) the influence on the context/environment (social, ethical, economic impacts) (Kerr et al., 2020; OECD, 2020). Such a wide scope of AI may be a consequence of the diversity of its development and applications. For instance, physicists and AI researchers work together to analyze and interpret “unmanageable volumes of data” produced by particle colliders (like the Large Hadron Collider) using specific models and algorithms to discover complex properties of matter (like the Higgs boson) (Castelvecchi, 2015, p. 18), and in medicine, the collaboration between physicians and AI researchers has led to the deployment of robots to assist remote surgery (Hamet & Tremblay, 2017) and nanorobots to precisely deliver drugs inside the human body (Hassanzadeh et al., 2019).

Machine learning (ML) has been the main tool for leveraging the complexity and usefulness of AI (Collins et al., 2021; Liu et al., 2021). ML evolved from simple pattern recognition through biological neuronal emulation (Rosenblatt, 1960; Uhr & Vossler, 1961) to recognition, identification, analysis, and classification of complex—and sometimes incomplete—contents, such as ancient texts (Assael et al., 2022). The profound developments in ML—nowadays under the umbrella term *deep learning*—have been carried out in countries with immense capacity to invest in research and infrastructure (Liu et al., 2021) with the support of research centers and universities (Shao et al., 2020).

However, while studies on AI investments have addressed the role of nations and governments, the role of companies has been anecdotal (e.g., Kovacevich, 2022; Newman, 2017; Webb, 2019). The state of the art in AI seems to have been defined by companies like Alphabet (Google), which develops TensorFlow, a powerful ecosystem for ML in use in many industries (Pang et al., 2020), and Microsoft/Open AI's chatGPT, a tool for processing natural language that has become the standard in applications from search queries to language translation (Edwards, 2021). As another example, Meta delivered its SEER self-supervised computer vision model that can learn without data curation and labeling (Ramanathan et al., 2021), both of which are well-known pre-processing tasks in conventional computer vision training. In such a business environment, some authors identify a "race to AI" (Smuha, 2021, p. 3) leading to a "winner-takes-all" phenomenon (Ding & Dafoe, 2021, p. 192).

Hypotheses

Dominance over an industry favors the over-concentration of power either through the accumulation of knowledge (Marimon & Quadrini, 2011) or due to ownership of the most efficient resources (Arnosti & Weinberg, 2022), thus imposing barriers to competition. By the same token, we can expect that cutting-edge AI development is made possible by companies that possess the most efficient hardware as well as data in greater volume and variety. Assuming that investments in R&D and in equipment serve as a proxy for AI investments (Yiu et al., 2020; Benetta et al., 2021), we hypothesize:

> H_1: R&D investments by companies lead to the formation of a group of AI leading companies (LC).

Due to complexity and research costs, companies turn to partnerships in R&D (Choi & Contractor, 2019), preferably with universities (Mahdi et al., 2019). Although many universities participate in R&D with the industry, we can expect that some universities stand out due to the amount of resources they have available to invest in R&D, specifically R&D focused on computer science as a proxy for AI expenditures. Therefore, we hypothesize:

> H_2: R&D expenditure by universities leads to the formation of a group of AI leading universities (LU).

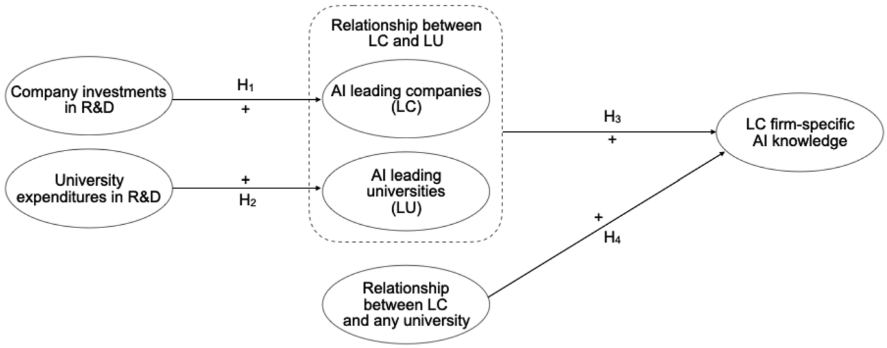

Fig. 1 Conceptual model

Companies that occupy a central position in an interfirm research cooperation achieve greater efficacy in knowledge acquisition and innovation, and "[they] will likely benefit more from diverse partners" (Wang et al., 2020, p. 171), including universities and research centers. Such a strategic positioning is better spotted by the industry than by universities (Shao et al., 2020). Therefore, we expect that universities focus on the production of factual knowledge, but they do not necessarily consider the integration of resources, whereas companies devise the big picture of an organizational cooperation and the intended outcomes. Therefore, we hypothesize:

H_3: The partnership between LC and LU leads to the production of LC firm-specific AI knowledge.

Finally, universities are not uniform in producing knowledge (Huggins et al., 2012), with some universities being more prone than others to engaging in collaborative partnership with business companies (Johnston, 2019). This makes us expect that not only LU collaborate with LC for the production of AI knowledge, but other universities also do it. Therefore, we hypothesize:

H_4: The partnership between LC and universities in general leads to the production of LC firm-specific AI knowledge.

Figure 1 synthesizes the idea that institutions investing the most in R&D are the ones producing the most in AI, and this is done through collaborative strategies (partnerships). The underlying assumption is that a few companies produce relevant AI with the support of universities, which in turn gives rise to an over-concentration of firm-specific AI knowledge and barriers to competition.

Method

We used three sources of secondary data for the test of hypotheses. The first source contains R&D investments from companies (dataset "A"), the second source contains R&D expenditures from universities (dataset "B"), and the third source contains the academic output (papers presented in AI conferences) of companies when they produce alone as well as the academic output of companies in partnership with universities (dataset “C”).

R&D investments have been adopted as a predictor for the operational capability in high-tech firms (e.g., Yiu et al., 2020), and a predictor for AI investments in the European context (e.g., Benetta et al., 2021). We followed the same approach for datasets A and B: dataset A comprises the European Commission’s industrial R&D investment scoreboard (EC-IRI) for the 2500 companies that invested the largest sums in R&D worldwide each year; and dataset B comprises the National Science Foundation’s Higher Education Research and Development (NSF-HERD) report on R&D expenditures from US universities, considering that US universities are among the best in all global performance rankings (Shanghai, THE, etc.). Both datasets A and B were accessed on March 2022, and they included data from 2012 to 2021 published each subsequent year by the EC-IRI and NSF-HERD reports.

The academic output (papers published in conference proceedings) has been adopted as a predictor for the relationship between companies and universities (e.g., Li et al., 2021). We followed this same approach to build dataset C from papers presented at conferences listed in the *2021-2022 International Conferences in Artificial Intelligence, Machine Learning, Computer Vision, Data Mining, Natural Language Processing and Robotics*[1] and provided that their h-index was at least 100. Six conference websites were accessed on August 2022, comprising data from 2013 to 2022: *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), *Neural Information Processing Systems* (NeurIPS), *International Conference on Computer Vision* (ICCV), *Annual Meeting of the Association for Computational Linguistics* (ACL), *Association for the Advancement of Artificial Intelligence* (AAAI), and *Conference on Empirical Methods in Natural Language Processing* (EMNLP).

As for the analysis of data, we adopted cluster analysis for H_1 and H_2, Web scraping and classification algorithms for H_3 and H_4, and the Mann-Whitney U for the test of hypotheses. Cluster analysis is a ML technique that does not require labeled data (non-supervised training) (Lai et al., 2019), which is the case for datasets A and B, and it offers a way to find patterns in raw data by grouping lines/observations with common characteristics (Li et al., 2020). Cluster analysis contributes to this study because it is suitable for grouping companies according to how closely associated their R&D investment capabilities are, which is required to answer the first research question.

Among the multiple classes of clustering algorithms available to test H1 and H2 (e.g., hierarchical, partition, grid, and density-based), we have opted for the density-based one, as it does not assume normality, does not depend on the shape of data, and deals well with noise (Lai et al., 2019). The density-based spatial clustering of applications with noise (DBSCAN; Martin et al., 2001) implements a density-based clustering and has been considered a standard in its class (Pelka, 2018; Fredström et al., 2021). DBSCAN does not require the *a priori* definition of the number of clusters to be found (Li et al., 2020), such as

[1] https://jackietseng.github.io/conference_call_for_paper/conferences-with-ccf.html

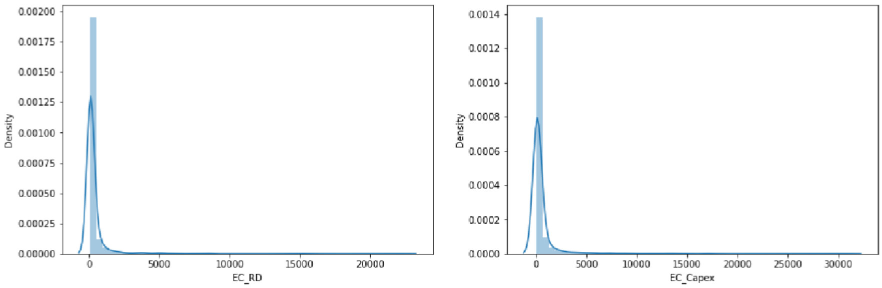

Fig. 2 Kurtosis analysis on R&D investments and Capex Source: developed by the authors

in the case of *k*-means. Instead, it requires two other parameters: Eps (the radius of a centroid) and MinPts (minimum amount of data points reachable within the centroid radius). Estimating both central parameters for the DBSCAN algorithm has been a challenge (Braune et al., 2015), to the point that many researchers try to automate such an estimation (e.g., Hou et al., 2016; Karami & Johansson, 2014), while others claim that this is a decision that is often based on a researcher's experience (Lai et al., 2019) through an iterative approach (Soni & Ganatra, 2016), sensitivity analysis (Fredström et al., 2021), or domain knowledge and heuristics (Schubert et al., 2017). Our approach followed Braune et al. (2015) and Soni and Ganatra (2016), assuming the outlier threshold as a determinant for both parameters.

The initial analysis of data suggested that the biggest investors are outliers (Fig. 2), which is a first important finding. Considering the presence of companies with massive investments in R&D and Capex (capital goods) in the data, we take each of those companies as a single cluster. Following such a rationale, MinPts=1 (at least one company/cluster) and Eps=2.5 (the outlier threshold radius—Euclidean distance—identified through iterative attempts). After finding the clusters, we applied Mann-Whitney tests to verify the statistical significance of differences as a non-parametric alternative to two independent sample *t*-tests.

For H_3 and H_4, we used Web scraping and specific classification algorithms. Web scraping is a technique to extract data directly from the World Wide Web using automated algorithms to convert the content of websites—usually, non-structured content—into structured datasets (Zhao, 2017). And the specific classification algorithms, in their turn, contribute to this study because we did not find any current technique to access and sort the conference papers and subsequently extract the precise information we needed to answer the second research question. We went through every Web page of AI conference proceedings in search of links to PDF files containing the papers. Then, we extracted the paper titles, authors, affiliations, and funding information (from the acknowledgments) to build dataset C. "Appendix" summarizes the procedures we have adopted for mining, collecting, and processing the three datasets, and it also explains how those sources were used to answer the research question.

Table 1 Top 20 largest R&D and Capex investors (2012–2021)

Company	Year									
	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012
Samsung electronics	1	4	2	1	3	2	4	6	5	5
Alphabet (Google)	2	1	1	6	6	11	13	26	46	44
Toyota Motor	3	3	3	2	1	1	2	8	6	4
Microsoft	4	5	8	7	11	13	16	22	25	28
PetroChina	5	2	4	3	5	3	1	1	1	2
Facebook	6	10	11	19	35	55	100	155	171	323
Volkswagen	7	7	7	5	4	6	8	10	9	11
China Mobile (2021)/China Telecom (2020-)	8	37	32	26	21	1533	27	41	57	53
Huawei investment & holding	9	13	15	16	20	42	64	93	103	104
Intel	10	9	9	8	10	14	12	15	12	12
Saudi Arabian Oil (Aramco)	11	6	5	-	-	-	-	-	-	-
General Motors	12	8	6	4	2	4	15	21	22	19
Apple	13	11	10	9	7	15	22	33	31	37
NTT	14	18	19	27	24	217	25	23	13	9
Taiwan Semiconductor	15	20	31	24	26	37	39	42	45	48
Kalera	16	-	-	-	-	-	-	-	-	-
China Petroleum & Chemical	17	16	24	40	36	18	14	12	10	8
Électricité de France	18	17	21	13	17	19	20	17	18	415
Exxon Mobil	19	12	14	15	14	7	5	7	3	2
Roche	20	23	23	21	22	24	35	39	37	31

Source: developed by the authors

Cells contain the ranking position of each company in the corresponding year. All companies in this list were founded before 2012. The newest one is Kalera, founded in 2010

Samples

Dataset A for the 2021 edition of the EC-IRI report includes 779 (31%) US companies, 597 (24%) Chinese companies, 401 (16%) EU companies, 293 (11%) Japanese companies, and 430 (17%) companies from the rest of the world. Dataset B for the 2021 edition of the NSF-HERD report includes 655 US universities. The numbers are nearly the same for all other editions of each report. Dataset A is formed by tuples with 19 items, including the company, the country, R&D, R&D 1-year-growth, and Capex (capital goods). Dataset B is formed by tuples with 13 items, including the university, all R&D expenditure, R&D expenditure on computer science, and R&D expenditure on information science. Both datasets were pre-processed (file name, column label, and data-type standardization) and analyzed for distribution, missing values, and outliers. Table 1 shows the top 20 R&D and Capex investors.

From dataset A, we chose R&D investments (EC_RD, henceforward) and Capex (EC_Capex, henceforward) as the items of interest. Missing values are usually handled by marginalization (discarding the missing values) or imputation (filling in missing values) (García-Laencina et al., 2010). The option for imputation implies choosing one of several strategies to obtain the value for each missing value. No missing values were found for EC_RD. Missing values found for EC_Capex (128 cases on average per year/report, or 5.1%) were filled with zero, considering that replacing it with, say, the mean or the median instead of zero would artificially amplify investments made in capital goods, with potential distortions on the clustering. Also, inference would not be suitable in this case because Capex does not depend on other factors in our dataset.

From dataset B, we chose all R&D expenditure (All_RD, henceforward) and R&D expenditure on computer and information sciences (CIS, henceforward) as the items of interest, considering that investments in R&D and CIS, especially made by tech companies, have been related to AI investments (Liu et al., 2021). No missing values were found for All_RD and CIS. Outliers were evaluated with the Mahalanobis distance (Riani et al., 2009) with a 0.001 significance level.

Dataset C includes 36,411 papers with data from AI conferences from 2013 to 2022. Dataset C is formed by tuples with four items: paper title, authors, affiliations, and funding. With an algorithm specifically developed for classification, each paper was classified according to four classes: (C1) authors' affiliation, with the following three categories: from universities, from companies, and from universities and companies; (C2) research funding, with the following seven categories: from universities, from companies, from government, from universities and companies, from universities and government, from companies and government, and from universities, government, and companies; (C3) a subclass of C1 when its content is "from university" in the form of a counter for the occurrence of each university according to Table 2; and (C4) a subclass of C1 when its content is "from company" in the form of a counter for the occurrence of each company according to Table 3. All algorithms developed for the classification of dataset C were implemented separately by two of the authors of this study, with a 98.9% inter-rater agreement. Table 4 shows the quantity of papers extracted for each conference between 2013 and 2022.

Results

Analysis of Datasets A and B and Test of Hypotheses H_1 and H_2

Figure 2 shows a leptokurtic shape for dataset A (which is also the case for dataset B). A leptokurtic shape "is a widely accepted" issue in financial data (Premaratne & Bera, 2005, p. 169). It includes Poisson, logistic, and student-*t* distributions, and it suggests the occurrence of outliers (Bossaerts, 2021). So, data distribution is abnormal and includes at least two data segments: data concentration and data dispersion (outliers). Further analysis on outliers shows that the current biggest R&D investors are companies like Alphabet, Huawei, Microsoft, Samsung, Apple, and Facebook.

Table 2 Dataset C (2013–2022)

Conference	h-index	Number of papers									
		2022	**2021**	**2020**	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**
CVPR: IEEE Conference on Computer Vision and Pattern Recognition	299	2074	1545	1466	1294	978	790	643	603	540	471
NeurIPS: Neural Information Processing Systems	198		2329	1905	1427	1009	679	569	403	411	360
ICCV: International Conference on Computer Vision	176		1582		1077		621		526		454
ACL: Annual Meeting of the Association for Computational Linguistics	135		572	794	753	533	246				
AAAI: Association for the Advancement of Artificial Intelligence	126	1269	1556	1558	1318						
EMNLP: Conference on Empirical Methods in Natural Language Processing	112		849	752	1302	1153					

Source: developed by the authors

Table 3 Clustering parameters and coefficients

Year	Dataset A			Dataset B		
	Clusters	SC	Eps	Clusters	SC	Eps
2012	9	.858	1.3	9	.826	1.1
2013	9	.881	1.68	9	.779	1.1
2014	9	.928	1.6	9	.784	.75
2015	9	.930	1.8	9	.775	.7
2016	9	.925	1.8	9	.779	1.2
2017	8	.932	1.7	8	.779	1.1
2018	9	.918	1.7	8	.762	1
2019	8	.936	2.5	9	.808	1.19
2020	9	.934	2.5	9	.792	1.19
2021	9	.936	2.5	8	.789	1.13

Source: developed by the authors

Figure 3 shows an example of cluster analysis applied to dataset A for the year 2021. The same analysis was done for each dataset and year. The scatter plot shows a region of high concentration at the bottom/left, and a region of dispersion at the top/right, with clusters in different colors.

The silhouette coefficient (SC) is a well-known measure of fitness for clustering (Pelka, 2018). SC assumes values from −1 to 1, and "[t]he higher value means the better assignment of objects into clusters" (Řezanková, 2018, p. 3). Table 5 shows the clustering parameters and coefficients for datasets A and B considering all years.

Table 4 Dataset A: clusters and companies (2021)

Cluster ID	Companies
#0	Alphabet
#1	Huawei Investments & Holding Apple
#2	Microsoft
#3	Samsung Electronics
#4	Facebook Volkswagen Intel
#5	Roche Johnson & Johnson Daimler + 4,485 other companies
#6	Toyota Motor
#7	PetroChina
#8	Saudi Arabian Oil China Mobile

Cluster ID is informed by the DBSCAN algorithm

Source: developed by the authors

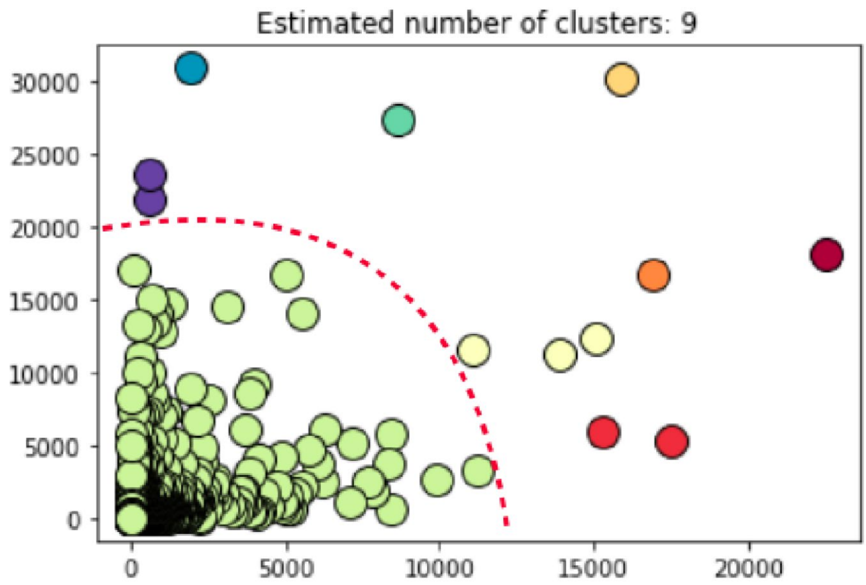

Fig. 3 Exemplary cluster analysis on dataset A (2021). Note: the dotted line suggests a threshold between ordinary investors and outliers Source: developed by the authors

Table 4 shows the clustering objects for dataset A, using data from 2021 to illustrate the clustering pattern we observed in all years. While 4488 companies form a single cluster (#5), 12 other companies form eight clusters.

Table 5 shows the clustering for dataset B, 2021. In this case, 646 universities form a single cluster, and nine other universities form eight additional clusters.

Table 5 Dataset B: clusters and universities (2021)

Cluster ID	Universities
#0	Johns Hopkins
#1	University of Michigan, Ann Arbor University of California, San Francisco University of Pennsylvania + 643 other universities
#2	University of Maryland, College Park
#3	Georgia Institute of Technology (GA Tech)
#4	Pennsylvania State University, University Park and Hershey Medical Center University of Illinois, Urbana-Champaign (UI)
#5	Massachusetts Institute of Technology (MIT) University of Southern California (USC)
#6	University of Texas, Austin (UT)
#7	Carnegie Mellon University (CMU)

Cluster ID is informed by the DBSCAN algorithm

Source: developed by the authors

Tables 6 and 7 respectively show the clustering from both datasets A and B in all years (2012–2021) obtained through the following steps, for dataset *k*, year *y*:

- First-order clustering using the algorithm DBSCAN has identified a number of c_{ky} clusters.
- Each cluster *C* is composed of *j* objects *o*, following the form $C_{k,y} = \{o_1, ..., o_{\mathrm{j}}\}$.
- Second-order clustering identified objects o_{j} often composing clusters c_{ky} (supposedly the ones investing heavily and permanently in R&D).

Table 6 shows the following:

1. There is a group of five companies that invest heavily and steadily in R&D and capital goods (Samsung Electronics, Intel, Toyota Motor, PetroChina, and Volkswagen). These are companies in the industries of "Software & Computer Services," "Electronic & Electrical Equipment," "Technology Hardware & Equipment," and "Automobiles & Parts."
2. There is a group of seven companies that have been gradually increasing investments in R&D and capital goods (Microsoft, Alphabet, Apple, Huawei, Aramco, Facebook, and China Mobile). These are companies in the "Software & Computer Services" and "Technology Hardware & Equipment" industries, with one company from the "Oil & Gas" industry.
3. There is a third group of 14 companies that have been gradually reducing investments in R&D and capital goods when compared to the other groups. This group consists mainly of companies in the "Oil & Gas," "Automobiles & Parts," and "Pharmaceuticals & Biotechnology" industries. The existence of these three second-order clusters suggests that there is an ongoing movement (in the larger R&D scenario) that can be explained by the occupation of empty spaces—companies occupying open investment spaces by companies from other industries (mostly "Oil & Gas")—or an increased pressure from companies that "push" from other industries in the race for AI.

Table 7 shows the following:

1. There is a group of eight universities that expend heavily and steadily on R&D in general and on R&D related to CIS in particular (Johns Hopkins, Georgia Tech, Pennsylvania State University, MIT, USC, UT, UI, and CMU).
2. There is a second group of seven universities that have been gradually reducing expenditure on R&D and CIS when compared to the other groups. And unlike companies, there is no group of late-entrant universities (i.e., universities that have been gradually increasing expenditure on R&D).

To test H_1, we searched for a significant difference in investments in R&D and equipment (Capex) between groups of companies that invest more in R&D. To test H_2, we searched for a significant difference in expenditure in R&D between groups of universities that expend more in R&D. To assess both hypotheses, we analyzed each

Table 6 Dataset A: clusters and companies (2012–2021)

Company	Year									
	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012
Samsung electronics	3	3	1	0	2	1	1	1	1	0
Intel	4	1	2	1	1	2	3	3	3	2
Toyota Motor	6	5	5	5	5	4	5	5	4	0
PetroChina	7	7	7	7	7	7	6	7	6	5
Volkswagen	4	1	2	1	0	0	0	0	0	2
Microsoft	2	1	2	1	1	2	2	2		
Alphabet (Google)	0	0	0	1	1	2	3			
Apple	1	1	2	1	4	5				
Huawei investment & holding	1	2	3	2						
Saudi Arabian Oil (Aramco)	8	8	7							
Facebook	4	1								
China Mobile	8									
General Motors		6	6	6	5	6				
Royal Dutch Shell				8	6	8	8	8	7	6
AT&T				8	6					
Daimler				4						
Exxon Mobil						8	8	8	7	7
Chevron						8	8	8	7	6
Gazprom						8		8	8	8
Total						8	7	8		
Petrobras							8	8		
Novartis								4		
Roche								4		
Nissan Motor									5	3
General Electric									5	3
NTT									5	4

Table 6 (continued)

Source: developed by the authors

Cells contain the ID of each cluster (ID informed by the DBSCAN algorithm). Colors represent “discretionary clusters” grouped by chronological assignment to DBSCAN clusters and quantity of assignments. “Year” is the year of data publication (fiscal year+1, where the fiscal year is a 1-year period that companies and governments use for reporting financial data)

year independently, thus forming 10 sub-hypotheses for each original hypothesis. Such a decision considers that objects o_j (companies and universities) often form distinct clusters c_{ky} with different R&D budget year after year. In order to form the groups for

Table 7 Dataset B: clusters and universities (2012–2021)

University	Year									
	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012
Johns Hopkins	0	0	0	0	0	0	0	0	0	0
Georgia Institute of Technology (GA Tech)	3	4	4	4	4	6	6	7	6	5
Pennsylvania State University, University Park and Hershey Medical Center	4	5	3	3	3	5	5	5	5	4
Massachusetts Institute of Technology (MIT)	5	3	2	2	2	4	4	5	4	4
University of Southern California (USC)	5	6	5	4	5	7	6	7	6	6
University of Texas, Austin (UT)	6	5	6	5	5	5	7	7	6	5
University of Illinois, Urbana-Champaign (UI)	4	7	7	4	5	6	6	6	6	5
Carnegie Mellon University (CMU)	7	8	8	7	7	8	8	8	8	8
University of Maryland, College Park	2	2		6	6				7	7
University of California, San Francisco						2	3	4	3	3
University of California, San Diego						3		3	2	
University of Michigan, Ann Arbor							1	1		1
University of Washington, Seattle									2	
Ohio State University									5	
University of Chicago										7

Source: developed by the authors

Cells contain the ID of each cluster (ID informed by the DBSCAN algorithm). Colors represent “discretionary clusters” grouped by chronological assignment to DBSCAN clusters and quantity of assignments. “Year” is the year of data publication (fiscal year+1, where the fiscal year is a 1-year period that companies and governments use for reporting financial data)

Table 8 Datasets A and B, Mann-Whitney *U* tests for H_1 and H_2 (2012–2021)

Year	Dataset A			Dataset B		
	H_1	Statistic	*p*-value	H_2	Statistic	*p*-value
2012	$H_{1,a}$	29,279.0	< .001	$H_{2,a}$	10,552.0	< .001
2013	$H_{1,b}$	29,491.0	< .001	$H_{2,b}$	8131.0	< .001
2014	$H_{1,c}$	27,030.0	< .001	$H_{2,c}$	6815.0	< .001
2015	$H_{1,d}$	31,978.0	< .001	$H_{2,d}$	6053.0	< .001
2016	$H_{1,e}$	27,074.0	< .001	$H_{2,e}$	6107.0	< .001
2017	$H_{1,f}$	33,833.0	< .001	$H_{2,f}$	5451.0	< .001
2018	$H_{1,g}$	29,150.0	< .001	$H_{2,g}$	5486.0	< .001
2019	$H_{1,h}$	33,868.0	< .001	$H_{2,h}$	4909.0	< .001
2020	$H_{1,i}$	23,092.0	< .001	$H_{2,i}$	5536.0	< .001
2021	$H_{1,j}$	17,146.5	< .001	$H_{2,j}$	5610.0	< .001

Source: developed by the authors

comparison (i.e., to assess the difference between two samples), groups were formed by objects that are above and below the threshold between ordinary investors and outliers, as illustrated in Fig. 3. The procedure was repeated for each dataset and each year. Table 8 shows that hypotheses H_1 and H_2 found statistical support.

Analysis of Dataset C and Test of Hypotheses H_3 and H_4

We have hypothesized that the role of universities in the production and exploitation of firm-specific knowledge is peripheral from the perspective of the final product design, production, and marketing, and it occurs because (1) companies seek profit and competitive advantage, whereas universities focus on research rather than attempting to market products and services; and (2) even when universities focus on entrepreneurial activities mainly through the creation of spin-off firms, such spin-offs are often bought by companies. So, to test hypotheses H_3 and H_4, we analyzed the academic output (from dataset C) of LC (clustered from dataset A) and LU (clustered from dataset B). Table 9 shows the quantity of papers published (the academic output) by authors affiliated to LC from 2013 to 2022.

From the 36,411 papers in dataset C (Table 9), 5742 (15.7%) have authors affiliated with LC. Of those, 2107 (36.7%) include a statement on funding, with 33.5% of funding coming from government and universities, which suggests that LC research funding is mostly internal (66.5%). Table 10 additionally shows that 711 papers were co-authored with LU. This seems to be not significant, but when we consider the collaboration with any university, we find 4161 papers. That is, 72.4% (about three-fourths) of papers published by the 12 LC are co-authored with universities.

To test H_3, we searched for a significant difference between the sum of papers authored by LC (last column of Table 9) and the quantity of co-authorships between LC and LU (penultimate column of Table 10). To test H_4, we searched for a significant difference between the sum of papers authored by LC (last column of Table 9) and the quantity of co-authorships between LC and any

Table 9 Database C: quantity of papers authored by LC (2013–2022)

Company	Number of papers										
	2022	**2021**	**2020**	**2019**	**2018**	**2017**	**2016**	**2015**	**2014**	**2013**	**Total**
Alphabet/Google	109	463	410	384	242	125	67	63	33	27	1,923
Microsoft	107	360	311	307	159	116	56	94	55	80	1,645
Apple	10	15	15	22	12	2	0	1	0	0	77
Huawei	108	217	115	84	23	2	0	2	8	5	564
Samsung	35	62	55	61	16	3	2	3	0	5	242
Facebook	33	257	174	229	88	42	12	18	4	3	860
Intel	18	44	33	64	33	28	12	8	3	7	250
Volkswagen	0	3	0	3	4	0	0	0	0	1	11
Toyota	15	22	20	38	18	10	11	5	4	13	156
Petro China	0	1	0	0	1	0	0	0	0	0	2
China Mobile	0	3	6	0	0	0	0	0	0	0	9
Saudi Arabian Oil/Aramco	0	2	0	1	0	0	0	0	0	0	3
Total	435	1449	1139	1193	596	328	160	194	107	141	5742

Source: developed by the authors

university (last column of Table 10). Firm-specific AI knowledge was measured as the quantity of papers published by authors affiliated to LC. To assess H_3, the statistical test considered the null hypothesis as the absence of any difference between the volume of academic production from the LC and the academic output from LC in collaboration with LU. Similarly, for H_4, the statistical test posited the null hypothesis as the absence of any difference between the volume of academic output from the LC and the academic output from LC in collaboration with universities in general. At a 95% confidence level, the application of the test indicated that we cannot reject the null hypotheses (respectively $p = 0.052$ and $p = 0.582$) in either scenario. Since the null hypotheses were not rejected in both instances, we have evidence to support the theoretical hypotheses H_3 and H_4, suggesting an intense partnership between LC and universities in general in producing firm-specific AI knowledge. Moreover, such an intense partnership between LC and universities in general is corroborated by a 72.4% rate of papers they have co-authored.

Table 10 also shows that companies more closely related to IT publish much more than the auto and oil ones at AI conferences. That is, R&D investments of the IT companies are positively correlated with the quantity of papers published in the main AI conferences. This is evidence in favor of using R&D investments as a proxy for AI investments, in the case of IT companies.

Table 10 Co-authorships between LC, LU, and any university (2013–2022)

Company	Number of papers co-authored between LC and LU										Number of papers co-authored between LC and any university
	JH	UM	GIT	PSU	UI	MIT	USC	UT	CM	Total	
Alphabet/Google	24	24	19	4	15	28	14	19	78	225	1204
Microsoft	14	27	12	4	38	8	4	40	71	218	1309
Apple	1	1	0	1	0	1	0	5	5	14	43
Huawei	11	3	0	4	1	1	1	3	3	27	509
Samsung	1	0	5	0	3	0	0	4	0	13	161
Facebook	21	13	35	1	14	4	8	25	53	174	656
Intel	0	4	3	0	3	1	0	4	6	21	173
Volkswagen	0	0	0	0	0	0	0	0	0	0	8
Toyota	0	0	0	0	4	1	1	2	10	18	87
Petro China	0	0	0	0	0	0	0	0	0	0	1
China Mobile	0	0	0	0	0	0	0	0	0	0	8
Saudi Arabian Oil/Aramco	0	0	0	0	0	0	0	1	0	1	2
Sum	72	72	74	14	78	44	28	103	226	711	4161

Source: developed by the authors

JH Johns Hopkins, *UM* University of Maryland, *GIT* Georgia Institute of Technology, *PSU* Pennsylvania State University, *UI* University of Illinois, *MIT* Massachusetts Institute of Technology, *USC* University of Southern California, *UT* University of Texas, *CM* Carnegie Mellon

Discussion

We proposed a rationale for the over-concentration of AI knowledge (OCAIK) as a consequence of the investment power of a few companies. In the proposal, the relationship between OCAIK and investments in R&D and capital goods is supported by the partnership with universities. The proposal foresees social, technical, ethical, and political impacts from such relationships. We consider that the model applies to contexts of competitiveness, in which companies that have slack resources seek to gain competitive advantages through the development of firm-specific knowledge.

Based on related works (Benetta et al., 2021; Yiu et al., 2020), we assume that investments by companies in R&D and capital goods (Capex) are suitable proxies for investments in AI. This seems to be a novel and useful approach. Even the number of patents or registration of new products is not a better variable to measure AI investments. For example, the AI Watch technical report from the European Commission Joint Research Centre uses OCDE data "to monitor the development, uptake and impact of AI for Europe" (Benetta et al., 2021, p. 1), but the results are an estimate as there are difficulties in computing patents and AI-related training (adopted as AI investment proxies). Our study, in its turn, found that companies more closely related to IT are much more involved in AI conference papers than companies from other industries. This is evidence in favor of using R&D as a proxy for AI in the case of IT companies. This study also differs from others in that we use institution-level data, whereas related work uses data aggregated by country, region, or industry.

We found a significant difference between clusters of companies that we now call *ordinary investors* and *outlier investors* (H_1). In 2021, for example, we found 4488 companies forming a single cluster (#5), while 12 other companies formed eight different clusters. This is the default pattern found in other years as well in dataset A. This finding supports the proposed rationale, especially in the relationship between OCAIK, R&D, and capital goods investments. This is also in line with Marimon and Quadrini (2011) in that the dominance over an industry favors the over-concentration of knowledge, as well as in line with Arnosti and Weinberg (2022) in that the ownership of the most efficient resources interposes barriers to competition.

We also found support for H_2, i.e., there is a significant difference in expenditure on R&D between groups of universities that expend more in R&D. In 2021, for example, we found 646 universities comprising a single cluster, while nine other universities form eight different clusters. This is the default pattern for other years as well in dataset B. This finding supports the proposed rationale, especially about the role of universities in the relationship between OCAIK, R&D, and capital goods investments, which is also in line with Choi and Contractor (2019) who state that due to complexity and research costs, companies turn to partnerships in R&D, and also in line with Mahdi et al. (2019), who posit that partnerships occur preferably with universities.

Regarding the role of the partnership of universities on the relationship between R&D and capital goods investment, and LC firm-specific AI knowledge, we did not find a significant difference between the academic output of LC integrated with LU and the academic output of LC integrated with universities in general (including LU), from which we conclude that there is an intense partnership between LC and universities in the production of firm-specific knowledge, which supports H_3 and H_4. In fact, about three-fourths (72.4%) of papers published by 12 LC are co-authored with universities.

Such findings suggest the existence of OCAIK represented by a few companies with a high capacity for investment in association with universities. OCAIK can be illustrated with a singularity metaphor, like in the case of a black hole, where density and gravity are so high that even light cannot escape the force of attraction at the point of no return. Our metaphor resembles a differentiation singularity: an insurmountable distance—at least during a certain period of time—between those who compete and those who are beyond the limits of competition. In this scenario, a group of players is considered singular when the sum of firm-specific knowledge already available by a firm added to its capacity to invest in new firm-specific knowledge development surpasses the sum of publicly available knowledge on that subject added to the capacity of other players to invest in firm-specific knowledge. Business competition may thus reach a point of no return if only incremental innovation is devised instead of radical innovation (Lindbloom, 1959). The expression below synthesizes the idea:

(e1) S_{it} if $(\mathrm{SKA}_{it} + \mathrm{NSK}_{it}) > (\sum \mathrm{PK} + \sum_1^n \mathrm{NSK})$
S, differentiation singularity
i, firm index
t, specific period of time
SKA, firm-specific knowledge already available
NSK, capacity for new firm-specific knowledge development
PK, publicly available knowledge on that subject
n, number of other players

Theoretical Implications

This study makes a number of contributions to the RBV and KBV literature. First, it points out an apparent limitation of both theories: how much advantage is needed to keep a sustainable advantage, while still keeping it competitive? To the best of our knowledge, both theories have no answer to that question. Second, the study proposes a metaphor for OCAIK, that of a differentiation singularity, to be considered in RBV and KBV. Metaphors and other cognitive resources are useful to illustrate situations and help identify key issues to be studied, designed, and managed (Jácome de Moura & Porto-Bellini, 2019). From the idea of a differentiation singularity, we may conceive the existence of a business oligopoly able to dictate the rules of the AI industry without being threatened by any real competitor. A third contribution of this study is to avoid investigating the performance of overconcentrated knowledge sectors in the light of RBV, KBV, or any other theory of the firm that assumes the existence of competition. That is, this study proposes a limit for theoretical explanations and distinguishes itself from recent systematic literature reviews on RBV and KBV that point out future trends in the development of theories (e.g., Pereira & Bamel, 2021) while not setting theoretical limits. This study additionally contributes to the literature on AI governance and regulation as it offers empirical evidence for the existence of OCAIK, the role of universities, and the potential consequences of the dominance of only a few industry actors over the creation and the evolution of AI knowledge.

Managerial Implications

Our findings show that companies that master the state-of-the-art in AI have distanced themselves from others beyond possible competition, since they are likely to have better resources, investment capacity, or motivation to work closely with scholarly research. Policy makers must be aware of this fact when planning ways to stimulate both the development and the regulation of AI, since the already established business capacity, organizational networks, and market presence may be far more difficult to manage than one can devise. Similarly, other companies should also be aware of that if they plan to compete in the AI arena through mere incremental innovation.

Moreover, our findings show that about three-fourths of papers published by 12 LC are co-authored with universities. This fact has two implications. First, managers of companies interested in AI-related R&D can benefit from an already open channel of interaction with universities. And second, policymakers and university managers must be cognizant of the resources being allocated to fund scholarly research, with such funding possibly interfering in business competition.

Limitations and Future Studies

The first limitation of this study refers to the databases included in the analyses. Even if we took measures to include all the relevant sources, there is always the possibility that important sources were ignored. Another limitation is that the coding procedure was highly dependent on the authors' discretion, even if including independent coding and formal assessments of agreement between two of the authors. Another possible limitation involves the use of R&D as a proxy for AI. This limitation is also present in the literature, since even the number of patents or registration of new products are no better variables to measure AI investments (Benetta et al., 2021). And another limitation is that we have restricted our analysis to data showing the presence of companies and universities in AI scholarly conferences. While we have employed quality criteria to select the sources of data, it is inevitable to think of data selection bias, once papers in conferences represent but a fraction of what is under development in the partnership of businesses and the academic institutions, not to say in the full AI landscape.

As for future studies, besides addressing some of the limitations above, we suggest that researchers could measure the dominance of the global AI industry through companies' investments specifically focused on AI development. Such a metric should be obtained through specific regulation that requires the disclosure of data on investments in AI, as AI regulation is a trendy topic in the global agenda.

Conclusion

This study discussed OCAIK as a manifestation of the dominance over the global AI industry by a small number of companies that are able to articulate with resourceful universities the development and the use of cutting edge, inimitable AI knowledge. Scholarly research is thus done in the periphery of business and, in planned or fortuitous manners, feeds business with research findings that grant those companies with increasing competitive advantages over new entrants. We verify the occurrence of OCAIK by processing thousands of papers presented in AI academic conferences in the last decade. Besides, we offer a novel insight about the RBV and KBV theoretical traditions, in that business competition may reach a point of no return if only incremental innovation is devised instead of radical innovation to break the chains of knowledge accumulation and technological implementation by a limited number of agents. By presenting evidence of a business oligopoly able to dictate the rules of the AI industry without being threatened by any real competitor, this study thus conveys political, economic, and social implications, since such evidence—i.e., obtained through close monitoring of companies' behavior—is the first step to counteract oligopolistic tendencies in the AI industry. We suggest monitoring the AI industry through a permanent panel/observatory, which could be championed by research institutions or organizations such as the OECD or WTO, for example.

Appendix. Synthesis of the Procedures for Mining, Collecting, and Processing the Three Datasets

Dataset	Link	Procedure
A	https://iri.jrc.ec.europa.eu/scoreboard	Step 1: access the main link Step 2: access specific link for each year Step 3: download the data (in CSV format) Step 4: remove empty columns in each file, as well as descriptive headers and totaling lines Step 5: standardize column/variable names Step 6: data processed through IDE Spider Python v3.9.7 using scipy and sklearn libraries
B	https://www.nsf.gov/statistics/srvyherd/#tabs-2	Step 1: access the main link Step 2: access specific link for each year Step 3: download the data (in CSV format) Step 4: remove empty columns in each file, as well as descriptive headers and totaling lines Step 5: standardize column/variable names Step 6: data processed through IDE Spider Python v3.9.7 using scipy and sklearn libraries

Dataset	Link	Procedure
C	Link provided according to each conference. Examples: https://openaccess.thecvf.com/CVPR2021 https://proceedings.neurips.cc/paper/2021 https://openaccess.thecvf.com/ICCV2021 https://aclanthology.org/volumes/2021.acl-long/ https://aaai.org/Library/AAAI/aaai21-issue02.php#4 https://aclanthology.org/volumes/2021.emnlp-main/	Step 1: web scraping to access each PDF file published by each conference Step 2: extract the paper titles, authors, affiliations, and funding information from each PDF file, composing a new CSV file Step 3: classify each record (in the new CSV file) according to: 3.1: go through CSV, line by line 3.2: check if affiliations have any company 3.3: check if CSV affiliations have any universities 3.4: check if funding has any company 3.5: check if funding has any university 3.6: check if funding has any government agency 3.7: if the paper has affiliations from universities, then C1 = 1 3.8: if the paper has affiliations from companies, then C1 = 2 3.9: if the paper has affiliations from universities and companies, then C1 = 3 3.10: if the paper has funding from universities, then C2 = 1 3.11: if the paper has funding from companies, then C2 = 2 3.12: if the paper has funding from the government, then C2 = 3 3.13: if the paper has funding from universities and companies, then C2 = 4 3.14: if the paper has funding from universities and government, then C2 = 5 3.15: if the paper has funding from companies and government, then C2 = 6 3.16: if the paper has funding from universities, government, and companies, then C2 = 7 Step 4: data processed through IDE Spider Python v3.9.7 using scipy library Step 5: 98.9% inter-rater agreement between two independent developers

Data Availability Data will be made available upon request. No artificial intelligence tool was used in this study.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Arnosti, N., & Weinberg, S. M. (2022). Bitcoin: A natural oligopoly. *Management Science, 68*(7), 4755–4771. https://doi.org/10.1287/mnsc.2021.4095

Assael, Y., Sommerschield, T., Shillingford, B., et al. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature, 603*(7900), 280–283. https://doi.org/10.1038/s41586-022-04448-z

Bahoo, S., Cucculelli, M., & Qamar, D. (2023). Artificial intelligence and corporate innovation: A review and research agenda. *Technological Forecasting & Social Change, 188*, 122264. https://doi.org/10.1016/j.techfore.2022.122264

Bai, X., & Li, J. (2020). The best configuration of collaborative knowledge innovation management from the perspective of artificial intelligence. *Knowledge Management Research & Practice*, 1–13. https://doi.org/10.1080/14778238.2020.1834886

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Benetta, A. D., Sobolewski, M., & Nepelski, D. (2021). *AI Watch: 2020 EU AI investments* (No. JRC126477). Joint Research Centre. https://doi.org/10.2760/017514

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly, 24*(1), 169–196. https://doi.org/10.2307/3250983

Boiko, K. (2021). R&D activity and firm performance: Mapping the field. *Management Review Quarterly, 71*(1), 1–37. https://doi.org/10.1007/s11301-021-00220-1

Bossaerts, P. (2021). How neurobiology elucidates the role of emotions in financial decision-making. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.697375

Braune, C., Besecke, S., & Kruse, R. (2015). Density based clustering: Alternatives to DBSCAN. In *Partitional Clustering Algorithms* (pp. 193–213). Springer, Cham. https://doi.org/10.1007/978-3-319-09259-1_6

Bruderer, H. (2016). The birth of artificial intelligence: First conference on artificial intelligence in paris in 1951? In: *IFIP International Conference on the History of Computing*, 181–185. https://doi.org/10.1007/978-3-319-49463-0_12

Bui, X. N., Nguyen, H., Choi, Y., Nguyen-Thoi, T., Zhou, J., & Dou, J. (2020). Prediction of slope failure in open-pit mines using a novel hybrid artificial intelligence model based on decision tree and evolution algorithm. *Scientific Reports, 10*(1), 1–17. https://doi.org/10.1038/s41598-020-66904-y

Cai, Y., Ramis Ferrer, B., & Luis Martinez Lastra, J. (2019). Building university-industry co-innovation networks in transnational innovation ecosystems: Towards a transdisciplinary approach of integrating social sciences and artificial intelligence. *Sustainability, 11*(17), 4633. https://doi.org/10.3390/su11174633

Castelvecchi, D. (2015). Artificial intelligence called in to tackle LHC data deluge. *Nature, 528*(7580), 18–19. https://doi.org/10.1038/528018a

Challa, H., Niu, N., & Johnson, R. (2020). Faulty requirements made valuable: On the role of data quality in deep learning. In *2020 IEEE Seventh International Workshop on Artificial Intelligence for Requirements Engineering (AIRE),* 61–69. https://doi.org/10.1109/AIRE51212.2020.00016

Choi, J., & Contractor, F. J. (2019). Improving the progress of research & development (R&D) projects by selecting an optimal alliance structure and partner type. *British Journal of Management, 30*(4), 791–809. https://doi.org/10.1111/1467-8551.12267

Cohen, J. F., & Olsen, K. (2013). The impacts of complementary information technology resources on the service-profit chain and competitive performance of South African hospitality firms. *International Journal of Hospitality Management, 34*, 245–254. https://doi.org/10.1016/j.ijhm.2013.04.005

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2021.102383

D'Este, P., & Perkmann, M. (2010). Why do academics engage with industry? The entrepreneurial university and individual motivations. *The Journal of Technology Transfer, 36*(3), 316–339. https://doi.org/10.1007/s10961-010-9153-z

Dafoe, A. (2018). *AI governance: A research agenda*. Oxford, UK, Future of Humanity Institute, University of Oxford.

Dahiya, R., Le, S., Ring, J. K., & Watson, K. (2021). Big data analytics and competitive advantage: The strategic role of firm-specific knowledge. *Journal of Strategy & Management*. https://doi.org/10.1108/jsma-08-2020-0203

Degravel, D. (2011). Managing organizational capabilities: The keystone step. *Journal of Strategy and Management, 4*(3), 251–274. https://doi.org/10.1108/17554251111152270

Ding, J., & Dafoe, A. (2021). The logic of strategic assets: From oil to AI. *Security Studies, 30*(2), 182–212. https://doi.org/10.1080/09636412.2021.1915583

Edwards, C. (2021). The best of NLP. *Communications of the ACM, 64*(4), 9–11. https://doi.org/10.1145/3449049

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21*(10–11), 1105–1121. https://doi.org/10.1002/1097-0266(200010/11)21:10/11%3c1105::AID-SMJ133%3e3.0.CO;2-E

Etzkowitz, H., & Leydesdorff, L. (1995). The triple helix—university-industry-government relations: A laboratory for knowledge-based economic development. *EASST Review*, *14*(1), 14–19. Available at SSRN: https://ssrn.com/abstract=2480085

Ferreira, J. J., Fayolle, A., Ratten, V., & Raposo, M. (2018). *Introduction: The role of entrepreneurial universities in society*. Edward Elgar Publishing. https://doi.org/10.4337/9781786432469.00005

Fink, L. (2011). How do IT capabilities create strategic value? Toward greater integration of insights from reductionistic and holistic approaches. *European Journal of Information Systems, 20*(1), 16–33. https://doi.org/10.1057/ejis.2010.53

Formica, P. (2002). Entrepreneurial universities: The value of education in encouraging entrepreneurship. *Industry & Higher Education, 16*(3), 167–175. https://doi.org/10.5367/000000002101296261

Fredström, A., Wincent, J., Sjödin, D., Oghazi, P., & Parida, V. (2021). Tracking innovation diffusion: AI analysis of large-scale patent data towards an agenda for further research. *Technological Forecasting & Social Change*. https://doi.org/10.1016/j.techfore.2020.120524

García-Laencina, P. J., Sancho-Gómez, J. L., & Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: A review. *Neural Computing & Applications, 19*(2), 263–282. https://doi.org/10.1007/s00521-009-0295-6

Ge, S., & Liu, X. (2022). The role of knowledge creation, absorption and acquisition in determining national competitive advantage. *Technovation, 112*, 102396. https://doi.org/10.1016/j.technovation.2021.102396

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17*(S2), 109–122. https://doi.org/10.1002/smj.4250171110

Grassano, N., Hernandez, H., Fako, P., Tuebke, A., Amoroso, S., Georgakaki, A., ... & Panzica, R. (2022). *The 2021 EU Industrial R&D Investment Scoreboard*, No. JRC127360. Joint Research Centre. https://doi.org/10.2760/559391

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism, 69*(Supplement), S36–S40. https://doi.org/10.1016/j.metabol.2017.01.011

Hassanzadeh, P., Atyabi, F., & Dinarvand, R. (2019). The significance of artificial intelligence in drug delivery system design. *Advanced Drug Delivery Reviews, 151*, 169–190. https://doi.org/10.1016/j.addr.2019.05.001

Henard, D. H., & McFadyen, M. A. (2006). *R&D knowledge is power. Research-Technology Management, 49*(3), 41–47. https://doi.org/10.1080/08956308.2006.11657377

Hou, J., Gao, H., & Li, X. (2016). Dsets-DBSCAN: A parameter-free clustering algorithm. *IEEE Transactions on Image Processing, 25*(7), 3182–3193. https://doi.org/10.1109/TIP.2016.2559803

Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., & Li, K. (2021). Artificial intelligence security: Threats and countermeasures. *ACM Computing Surveys (CSUR), 55*(1), 1–36. https://doi.org/10.1145/3487890

Huggins, R., Johnston, A., & Stride, C. (2012). Knowledge networks and universities: Locational and organisational aspects of knowledge transfer interactions. *Entrepreneurship & Regional Development, 24*(7–8), 475–502. https://doi.org/10.1080/08985626.2011.618192

Jácome de Moura, P., Jr., & Porto-Bellini, C. G. (2019). Shared flow in teams: Team vibration as emergent property, metaphor, and surrogate measure. *Team Performance Management, 25*(7/8), 440–456. https://doi.org/10.1108/TPM-12-2018-0072

Jeon, S., Chang, Y. S., & Jo, S. J. (2024). Speed of catch-up and convergence of the artificial intelligence divide: AI investment, robotic, start-ups, and patents. *Journal of Global Information Technology Management*, *27*(1). https://doi.org/10.1080/1097198X.2023.2297636

Johnston, A. (2019). The roles of universities in knowledge-based urban development: A critical review. *International Journal of Knowledge-Based Development, 10*(3), 213–231. https://doi.org/10.1504/IJKBD.2019.103205

Karami, A., & Johansson, R. (2014). Choosing DBSCAN parameters automatically using differentiation evolution. *International Journal of Computer Applications, 91*(7), 1–11.

Kaur, D., Uslu, S., & Durresi, A. (2020). Requirements for trustworthy artificial intelligence—a review. In *International Conference on Network-Based Information Systems* (pp. 105–115). https://doi.org/10.1007/978-3-030-57811-4_11

Kerr, A., Barry, M., & Kelleher, J. D. (2020). Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951720915

Kovacevich, A. (2022). *Take it from a software engineer: Big tech's monopoly is stifling innovation*. Accessed June 28, 2022. https://www.newsweek.com/take-it-software-engineer-big-techs-monopoly-stifling-innovation-opinion-1718646

Lai, W., Zhou, M., Hu, F., Bian, K., & Song, Q. (2019). A new DBSCAN parameters determination method based on improved MVO. *IEEE Access, 7*, 104085–104095. https://doi.org/10.1109/ACCESS.2019.2931334

Li, K., Zhang, J., & Li, D. (2021). Research status and hotspot analysis of literature metrology in artificial intelligence field. *Journal of Physics: Conference Series, 2024*(1), 012055. https://doi.org/10.1088/1742-6596/2024/1/012055

Li, Z., Li, Y., Lu, W., & Huang, J. (2020). Crowdsourcing logistics pricing optimization model based on DBSCAN clustering algorithm. *IEEE Access, 8*, 92615–92626. https://doi.org/10.1109/ACCESS.2020.2995063

Lindbloom, C. E. (1959). The science of "muddling through." *Public Administration Review, 19*(2), 79–88.

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: Constructing and applying a new search strategy. *Scientometrics, 126*(4), 3153–3192. https://doi.org/10.1007/s11192-021-03868-4

Mahdi, O. R., Nassar, I. A., & Almsafir, M. K. (2019). Knowledge management processes and sustainable competitive advantage: An empirical examination in private universities. *Journal of Business Research, 94*, 320–334. https://doi.org/10.1016/j.jbusres.2018.02.013

Marakova, V., Wolak-Tuzimek, A., & Tuckova, Z. (2021). Corporate social responsibility as a source of competitive advantage in large enterprises. *Journal of Competitiveness, 13*(1), 113–128. https://doi.org/10.7441/joc.2021.01.07

Marimon, R., & Quadrini, V. (2011). Competition, human capital and income inequality with limited commitment. *Journal of Economic Theory, 146*(3), 976–1008. https://doi.org/10.1016/j.jet.2011.01.001

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings 8th IEEE International Conference on Computer Vision. ICCV 2001*, 416–423. https://doi.org/10.1109/ICCV.2001.937655

Mayer, K. J., Somaya, D., & Williamson, I. O. (2012). Firm-specific, industry-specific, and occupational human capital and the sourcing of knowledge work. *Organization Science, 23*(5), 1311–1329. https://doi.org/10.1287/orsc.1110.0722

Menke, M. M. (1997). Managing R&D for competitive advantage. *Research-Technology Management, 40*(6), 40–42. https://doi.org/10.1080/08956308.1997.11671169

Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128). https://doi.org/10.1098/rsta.2017.0357

Miotti, L., & Sachwald, F. (2003). Co-operative R&D: Why and with whom? An integrated framework of analysis. *Research Policy, 32*(8), 1481–1499. https://doi.org/10.1016/S0048-7333(02)00159-2

Miotto, G., Del-Castillo-Feito, C., & Blanco-González, A. (2020). Reputation and legitimacy: Key factors for higher education institutions' sustained competitive advantage. *Journal of Business Research, 112*, 342–353. https://doi.org/10.1016/j.jbusres.2019.11.076

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill Science/Engineering.

Mongeau, S., Hajdasinski, A. (2021). Managerial recommendations. In: *Cybersecurity Data Science*. Springer, Cham. https://doi.org/10.1007/978-3-030-74896-8_6

Nayak, B., Bhattacharyya, S. S., & Krishnamoorthy, B. (2022). Exploring the black box of competitive advantage—an integrated bibliometric and chronological literature review approach. *Journal of Business Research, 139*, 964–982. https://doi.org/10.1016/j.jbusres.2021.10.047

Newbert, S. L. (2008). Value, rareness, competitive advantage, and performance: A conceptual-level empirical investigation of the resource-based view of the firm. *Strategic Management Journal, 29*(7), 745–768. https://doi.org/10.1002/smj.686

Newman, D. (2017). *Inside look: The world's largest tech companies are making massive AI investments*. Accessed May 4, 2022. https://www.forbes.com/sites/danielnewman/2017/01/17/inside-look-the-worlds-largest-tech-companies-are-making-massive-ai-investments/?sh=4d7f17cc4af2

OECD. (2020). *A first look at the OECD's framework for the classification of AI systems, designed to give policymakers clarity*. Accessed May 2, 2022. https://oecd.ai/en/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers

Orlando, B., Ballestra, L. V., Magni, D., & Ciampi, F. (2021). Open innovation and patenting activity in health care. *Journal of Intellectual Capital, 22*(2), 384–402. https://doi.org/10.1108/JIC-03-2020-0076

Pang, B., Nijkamp, E., & Wu, Y. N. (2020). Deep learning with TensorFlow: A review. *Journal of Educational & Behavioral Statistics, 45*(2), 227–248. https://doi.org/10.3102/1076998619872761

Parmigiani, F., Haller, I., Gkantsidis, C., & Ballani, H. (2021, May). Optics for the cloud: Challenges and opportunities. In: *CLEO: Science and Innovations* (pp. STu1J-2). Optical Society of America.

Pełka, M. (2018). Analysis of innovations in the European Union via ensemble symbolic density clustering. *Ekonometria*, *22*(3). https://doi.org/10.15611/eada.2018.3.06

Pereira, V., & Bamel, U. (2021). Extending the resource and knowledge-based view: A critical analysis into its theoretical evolution and future research directions. *Journal of Business Research, 132*, 557–570. https://doi.org/10.1016/j.jbusres.2021.04.021

Premaratne, G., & Bera, A. (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics, 3*(2), 169–187. https://doi.org/10.1093/jjfinec/nbi009

Ramanathan, V., Wang, R., & Mahajan, D. (2021). Predet: Large-scale weakly supervised pre-training for detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2865–2875. https://doi.org/10.1109/ICCV48922.2021.00286

Řezanková, H. A. N. A. (2018). Different approaches to the silhouette coefficient calculation in cluster evaluation. In: *21st International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics*, 1–10.

Riani, M., Atkinson, A. C., & Cerioli, A. (2009). Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society: Series B, 71*(2), 447–466. https://doi.org/10.1111/j.1467-9868.2008.00692.x

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & society, 36*(1), 59–77. https://doi.org/10.1007/s00146-020-00992-2

Rosenblatt, F. (1960). Perceptron simulation experiments. *Proceedings of the IRE, 48*(3), 301–309. https://doi.org/10.1109/JRPROC.1960.287598

Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research & Development., 44*(1), 206–226. https://doi.org/10.1147/rd.33.0210

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS), 42*(3), 1–21. https://doi.org/10.1145/3068335

Shannon, C. E. (1950). Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 41*(314), 256–275. https://doi.org/10.1080/14786445008521796

Shao, Z., Yuan, S., & Wang, Y. (2020). Institutional collaboration and competition in artificial intelligence. *IEEE Access, 8*, 69734–69741. https://doi.org/10.1109/ACCESS.2020.2986383

Smuha, N. A. (2021). From a 'race to AI'to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation & Technology, 13*(1), 57–84. https://doi.org/10.1080/17579961.2021.1898300

Soni, N., & Ganatra, A. (2016). Aged (automatic generation of eps for DBSCAN). *International Journal of Computer Science & Information Security, 14*(5), 536.

Statista. (2022). Estimated number of companies worldwide from 2000 to 2020. *Statista*. Accessed March 28, 2022. https://www.statista.com/statistics/1260686/global-companies/#statisticContainer

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3c509::AID-SMJ882%3e3.0.CO;2-Z

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Uhr, L., & Vossler, C. (1961). A pattern recognition program that generates, evaluates, and adjusts its own operators. In: *Papers presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference*, 555–569. https://doi.org/10.1145/1460690.1460751

Ullah, Z., & Arslan, A. (2022). R&D contribution to sustainable product attributes development: The complementarity of human capital. *Sustainable Development, 30*(5), 902–915. https://doi.org/10.1002/sd.2289

Wang, C., Chen, M. N., & Chang, C. H. (2020). The double-edged effect of knowledge search on innovation generations. *European Journal of Innovation Management, 23*(1), 156–176. https://doi.org/10.1108/EJIM-04-2018-0072

Wang, H., Choi, J., Wan, G., & Dong, J. Q. (2016). Slack resources and the rent-generating potential of firm-specific knowledge. *Journal of Management, 42*(2), 500–523. https://doi.org/10.1177/0149206313484519

Webb, A. (2019). *The big nine: How the tech titans and their thinking machines could warp humanity*. NY, PublicAffairs.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://doi.org/10.1145/365153.365168

Yiu, L. D., Yeung, A. C., & Jong, A. P. (2020). Business intelligence systems and operational capability: An empirical analysis of high-tech sectors. *Industrial Management & Data Systems, 120*(6), 1195–1215. https://doi.org/10.1108/IMDS-12-2019-0659

Zahra, S. A., Neubaum, D. O., & Hayton, J. (2020). What do we know about knowledge integration: Fusing micro-and macro-organizational perspectives. *Academy of Management Annals, 14*(1), 160–194. https://doi.org/10.5465/annals.2017.0093

Zhao, B. (2017). Web scraping. *Encyclopedia of big data*, 1–3.

Authors and Affiliations

Pedro Jácome de Moura Jr.[1] · Carlos Denner dos Santos Junior[2] · Carlo Gabriel Porto-Bellini[1] · José Jorge Lima Dias Junior[1]

✉ Pedro Jácome de Moura Jr.
pjacome@sti.ufpb.br

Carlos Denner dos Santos Junior
carlosdenner@unb.br

Carlo Gabriel Porto-Bellini
cgpbellini@ccsa.ufpb.br

José Jorge Lima Dias Junior
jorge.dias@academico.ufpb.br

[1] Department of Management, UFPB, João Pessoa, Brazil

[2] Department of Management, UnB, Brasilia, Brazil