Technovation 146 (2025) 103295

Contents lists available at ScienceDirect

Technovation

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

What “V” of the big data support firms’ radical and incremental innovation?

Giulio Ferrigno *, Saverio Barabuffi, Enrico Marcazzan, Andrea Piccaluga

Scuola Superiore Sant’Anna, Institute of Management and L’EMbeDS Department, Piazza Martiri della Libertà, 33, Pisa, Italy

A R T I C L E I N F O

Keywords:
Big data
Radical innovation
Incremental innovation
Italy

A B S T R A C T

Despite the considerable attention from both academics and practitioners to the effects of big data on firms’ innovation performance, a noticeable research gap remains in understanding how big data influences different types of innovation—namely, radical and incremental innovation. Many studies recognize that big data can be a valuable source of innovation, as it enables firms to gather and incorporate insights from customers, partners, suppliers, and other stakeholders. However, prior research has rarely investigated this relationship through a granular lens, failing to distinguish the specific effects of big data on radical and incremental innovation.

Focusing on firms’ intent of introducing radical and incremental innovation using big data, we employ the Knowledge Based View and the four well-known dimensions of big data (i.e., volume, velocity, variety, and veracity) to explore if and when big data is a source of knowledge for radical and incremental innovation. Performing an OLS regression analysis on a sample of 155 Italian firms, we find that both big data variety and veracity positively affect firms’ radical and incremental innovation. These findings provide insights about the conditions under which big data can improve firms’ innovation processes, contributing to a more comprehensive theoretical understanding of the opportunities big data bring in the context of firms’ product, service and process innovation. Moreover, our findings offer valuable guidance to managers navigating the complexities of leveraging big data for new product development.

1. Introduction

Big data is one of the key drivers fostering firms’ innovation performance (Cappa et al., 2021; Ghasemaghaei and Calic, 2020; Nafizah et al., 2024; Salomo et al., 2008). The increasing digitalization of society in general and the business environment in particular stimulates the generation of large amounts of data on both the demand and the supply side (de Camargo et al., 2018; Ritala et al., 2024). Analyzing such amount of data is crucial to understand the behavior of a number of firms’ stakeholders (especially customers) (Bharadwaj and Noble, 2017; Chen et al., 2012) and to develop new products that satisfy their needs (Matthias et al., 2015; Zhang et al., 2015) or even anticipate them. In this vein, research has shown that four characteristics of big data - also known as 4Vs (Ghasemaghaei and Calic, 2019) - may reveal important insights for firms’ innovation performance. The 4Vs of big data are big data *volume* (i.e., the amount of data collected); *variety* (i.e., the assortment of data collected per observation); *velocity* (i.e., the speed of generating and analyzing data), and *veracity* (i.e., the reliability and insightfulness of data) (Cappa et al., 2021).

Accordingly, several scholars claim that big data characteristics do in fact represent an important source for firm’s new product development (hereafter NPD) process (Davenport and Stephan, 2016; Davenport et al., 2012; McAfee and Brynjolfsson, 2012) because firms can benefit from analyzing large amounts of data to develop new products and services (Bstieler et al., 2018; Johnson et al., 2017; Tan and Zhan, 2017). In this sense, as stated by Johnson et al. (2017), “one element that may improve this NPD process is the extent to which firms incorporate insights from customer knowledge (e.g., big data)” (Johnson et al., 2017, p. 8). Thus, the adoption of big data can increase the quality of data-driven decisions and, in turn, be considered as a driver of new product development process (Johnson et al., 2017).

So far, prior works that examined the impact of big data on firms’ NPD have mostly privileged qualitative methodologies (Capurro et al., 2021; Jagtap and Nguyen Khanh Duong, 2019; Tan and Zhan, 2017) or approached this relationship by analyzing the effects of big data on firms’ innovation efficacy vs efficiency (Ghasemaghaei and Calic, 2020)

* Corresponding author
E-mail addresses: giulio.ferrigno@santannapisa.it (G. Ferrigno), saverio.barabuffi@santannapisa.it (S. Barabuffi), enrico.marcazzan@gmail.com (E. Marcazzan), andrea.piccaluga@santannapisa.it (A. Piccaluga).

https://doi.org/10.1016/j.technovation.2025.103295
Received 23 December 2024; Received in revised form 7 May 2025; Accepted 9 June 2025
Available online 13 June 2025

or the effects of firms' exploration and exploitation orientation on big data usage (Johnson et al., 2017). Surprisingly, no previous study has tried to analyze the effects of big data on radical and incremental innovation[1] (Bstieler et al., 2018; Cappa et al. 2021; Mariani et al. 2023), which is the object of our paper. The theoretical importance of this issue is highlighted in the work by Mariani et al. (2023, p. 21), who claimed that a research area that deserves further investigation is to understand "what are the features of big data (volume, variety, velocity, veracity, etc.) that are most likely to support incremental and radical innovation". In line with this discourse, Cappa et al. (2021) suggest that "the effects of big data on firm performance could be further analyzed by considering firm-specific factors, such as the types of innovations delivered to the market, for example, incremental versus radical" (Cappa et al., 2021, p. 62; based on the work of Cammarano et al., 2019). Similarly, Bstieler et al. (2018) indicate that "there is still a lack of understanding of whether and how big data drives radical and disruptive innovation" (Bstieler et al., 2018, p. 304). Taken together, these studies suggest that an adequate understanding of big data-firm's NPD process requires a careful examination of the impact of big data characteristics on firms' radical and incremental innovation. In this paper, we therefore respond to these recent calls (Bstieler et al., 2018; Cappa et al., 2021; Mariani et al., 2023). From a managerial perspective, understanding which specific features of big data (e.g., volume, variety, velocity and veracity) most effectively support firms' incremental and radical innovation is crucial to strategically leverage "smart" data in new product development, optimize resource allocation, and enhance firms' market positioning and innovation performance. As such, the research question of our paper is "what are the features of big data (volume, variety, velocity, veracity, etc.) that are most likely to support incremental and radical innovation"?

To tackle this research question, we ground our study on the Knowledge-Based View (KBV; Grant, 1996) and propose a theoretical framework that allows to move beyond a holistic approach (Mikalef et al., 2019). We offer a more nuanced examination of the conditions under which big data can be advantageous for firms' radical and incremental innovation. More specifically, drawing on this theory, we argue that big data characteristics can be a source of knowledge when firms are prone to introduce radical or incremental innovation in the market (Cooper et al., 2023). We hypothesize the effects of big data volume, variety, velocity and variety on firms' radical and incremental innovation. To verify these hypotheses, we perform an OLS regression analysis on a sample of 155 Italian firms. We decided to focus on Italy given the significant transition of its firms towards Industry 4.0 (Martinelli et al. 2021), also facilitated by government initiatives over the past decade, making it an intriguing landscape to study firms' adoption of big data (Petruzzelli et al., 2022).

On the whole, some of the hypotheses have been confirmed from OLS regression results and suggest that big data contribute to firms' radical innovation only in some specific cases. More precisely, our findings indicate that both variety and veracity positively affect firms' radical innovation. Interestingly, the same features also contribute to incremental innovation. These results allow us to provide a deeper theoretical understanding of the effects of big data on firms' innovation performance (Johnson et al., 2017; Mikalef et al., 2019; Radicic and Petković, 2023) and therefore provide an answer to unexplored research gaps in the innovation management literature (Bstieler et al., 2018; Cappa et al., 2021; Mariani et al., 2023).

The paper proceeds as follows. In the next section, we review the literature on big data characteristics and NPD and use the KBV to develop hypotheses about the effects of the 4 V of big data on firms' radical and incremental innovation. Then, we present the research methodology, the OLS regression analysis, and the results. After this, we discuss the findings of the study in the big data and NPD literature, reporting both theoretical and managerial contributions of our study. Finally, we describe some limitations of our study and related suggestions for further research.

2. Theoretical framework

2.1. Big data as a source of knowledge for firms' innovation: beyond a holistic view in favor of the "4Vs"

Advancements in information technology, the rise of digitalization and the development of social media, coupled with the diffusion of many devices that contain sensors and applications, can be considered as the main drivers of the "big data" phenomenon (Cappa et al., 2021). According to Del Vecchio et al. (2018), big data "refers to any set of data that, with traditional systems, would require large capabilities in terms of storage space and time to be analyzed" (Del Vecchio et al., 2018, p. 6). The availability of this kind of data can be a useful source of competitive advantage for firms (Acciarini et al., 2023; Del Vecchio et al., 2018), both in terms of innovation in general, and also – perhaps – in terms of radical innovation (Mikalef et al., 2019).

From a KBV perspective (Cooper et al. 2023; Grant, 1996), this means that big data can be considered as a valuable source of knowledge that allows firms to innovate and develop new products and services. In this sense, big data can be seen as crucial intangible assets for NPD (de Camargo et al., 2018). Initially conceptualized by the work of Cooper (1983) and Cooper and Kleinschmidt (1986), with subsequent revisions (see Cooper, 1990; Cooper, 2018), the NPD process can be conceived as a series of phases that start from idea generation and initial screening to the launch of a new product (Huang et al., 2002). In particular, Cooper (2018) provides insights regarding the importance of customers' knowledge in NPD process which can be elaborated upon the adoption and development of big data analytics capabilities and techniques (Tan and Zhan, 2017). In particular, within the NPD process, a crucial role is played by the choice of generating *incremental* or *radical* innovation (Eling et al., 2016; Marzi et al., 2021), respectively defined by Mikalef et al. (2019, p. 279) as "minor changes and modification to products and services" and "major departures from existing capabilities in the firm, which constitute the basis for completely new products, services or business models".

With regard to KBV, its application in the big data literature is extensive (de Camargo et al., 2018; Dwivedi et al., 2011). In fact, this theory has been playing a crucial role in stimulating discussions about the knowledge developed from handling large volumes of data, encompassing tasks such as filtering, analysis, and other elaborate actions (de Camargo et al., 2018). KBV also addresses the methods of generating knowledge and value from intangible assets such as big data, thereby contributing to the creation of value and a competitive advantage for organizations (Cooper et al., 2023; Dwivedi et al., 2011). Following this theoretical reasoning, several studies suggest that big data can be considered as a valuable source of knowledge for NPD (Davenport and Stephan, 2016; Johnson et al., 2017; Mikalef et al., 2019). In this regard, Johnson et al. (2017) argument that "one element that may improve this NPD process is the extent to which firms incorporate insights from customer knowledge (e.g., big data)" (Johnson et al., 2017, p. 8).

Adopting the KBV (Cooper et al., 2023; Grant, 1996), we therefore argue that big data can have an impact on firms' NPD processes, in particular in the introduction of radical and incremental innovation. In this regard, Mikalef et al. (2019) found that big data analytics capability, which is defined as the ability of a firm to capture and analyze data towards the generation of insights by effectively orchestrating and deploying its data, technology and talent (Gupta and George, 2016;

[1] Radical innovations represent major departures from existing capabilities in the firm, and constitute the basis for completely new products, services or business models (Ritala and Hurmelinna-Laukkanen, 2013). Instead, incremental innovations refer to minor changes and modification to products and services (Mikalef et al. 2019: p. 279).

Mikalef et al., 2018), empowers firms to derive insights that enhance their dynamic capabilities. This, in turn, has a positive effect on firms' radical innovation (Mikalef et al., 2019). Despite this insightful indirect correlation, the authors do not examine the specifics of how big data analytics capability precisely serves as a catalyst for a firm's radical and incremental innovation, thus leaving room for deeper examination.

To offer a more nuanced comprehension of big data as a source of knowledge for NPD and its impact on firm's radical and incremental innovation, we focus on its core characteristics, also known as "4Vs" (Acciarini et al., 2023; Cappa et al., 2021; Ghasemaghaei, 2021; Ghasemaghaei and Calic, 2019, 2020): *volume* (i.e., the amount of data collected); *variety* (i.e., the assortment of data collected per observation); *velocity* (i.e., the speed of generating and analyzing data) (Ghasemaghaei and Calic, 2020; Johnson et al., 2017); and *veracity (i.e., the reliability and insightfulness of data) (*Cappa et al., 2021; Ghasemaghaei and Calic, 2020). [2]

2.2. The "4Vs" of big data and firms' radical and incremental innovation

Johnson et al. (2017) investigated the intricate relationship between big data dimensions and new revenues. Their investigation uncovered a positive correlation between a firm's exploration orientation and the big data's dimensions of volume, variety, and velocity. However, intriguingly, the study revealed that the firm's exploitation orientation exhibited no significant association with the "3Vs" of big data. On the base of this nuanced – exploration-oriented – finding, we might suggest that volume, variety, and velocity play a more pronounced role in the development of radical innovation, rather than in the case of incremental innovation. As a matter of fact, this interpretation is aligned with the notion that a firm's exploration orientation, characterized by the quest for and the acquisition of new knowledge and skills, fosters an environment conducive to stimulating radical innovations (Andriopoulos and Lewis, 2009; Atuahene-Gima, 2005; Zhang et al., 2015).

Contrary to this perspective, Mikalef et al. (2019) do not posit a differential impact of big data on radical or incremental innovation. This theoretical ambiguity prompts us to consider both outcomes as potentially associated with the "4Vs" of big data. In the following sub-sections we formulate and expound upon the research hypotheses, which are summarized in Fig. 1.

2.2.1. The "4Vs" of big data and radical innovation

Relying extensively on vast amounts of data can offer numerous advantages to a firm, such as facilitating market capturing and enhancing the understanding of customer needs (Du and Kamakura, 2012). However, the assertion that "big data is not always better data" challenges such a notion (Ghasemaghaei and Calic, 2020, p. 151). Building on previous research, we contend that the volume dimension of big data may particularly undermine firms' radical innovation (Cappa et al., 2021). First, in the customer domain of the NPD process (Tan and Zhan, 2017), large datasets often fail to capture breakthrough insights needed for radical innovation, as they tend to reflect mainstream preferences and contain biases linked to specific social groups (Chang, 2021). Second, the cognitive constraints of human resources become particularly acute when dealing with radical innovation, as managers struggle to identify which unconventional data patterns merit attention amidst vast data volumes (Cappa et al., 2021; Prescott, 2016). Consequently, the substantial volume of big data may detrimentally affect idea generation quality, thereby reducing the likelihood of adopting radical innovation. In summary, while an ample volume of data can bring benefits, an excess of data could impede firms from fostering radical innovation. Therefore, we propose the following hypothesis:

H1. Big data volume has a negative effect on firms' radical innovation

In contrast to the first dimension, the dimension of data variety emerges as a particularly potent enabler of radical innovation (Cappa et al., 2021; Ghasemaghaei and Calic, 2019). More specifically, the integration of both structured and unstructured data sources provides firms with heterogeneous knowledge inputs that are critical for identifying disruptive opportunities (Cappa et al., 2021; Dean et al., 2023). This diversity allows firms to detect latent customer needs and atypical market patterns that conventional data sources might overlook, directly enhancing the novelty and originality of generated ideas - two hallmarks of radical innovation. By combining multiple data perspectives, firms can overcome the "dominant logic" trap that often constrains breakthrough thinking (Johnson et al., 2017). We therefore hypothesize:

H2. Big data variety has a positive effect on firms' radical innovation.

In line with big data variety, the dimension of data velocity appears to contribute positively to firms' radical innovation outcomes (Ghasemaghaei and Calic, 2019) by enabling real-time adaptation during the NPD process (Rakshit et al., 2021). In fast-moving environments where customer preferences evolve rapidly, the ability to process and act on fresh data streams allows firms to: (1) identify emerging needs before competitors (Ahmad et al., 2013), (2) rapidly prototype and test disruptive concepts (Dobbs et al., 2015), and (3) shorten the learning cycles essential for radical innovation (Ghasemaghaei and Calic, 2020). This temporal advantage is particularly crucial for radical innovations, which often target nascent or shifting market segments (Johnson et al., 2017). In line with the above arguments, we hypothesize the following relationship:

H3. Big data velocity has a positive effect on firms' radical innovation.

In harmony with the big data dimensions of variety and velocity, the dimension of veracity emerges as a crucial determinant with a positive impact on firm radical innovation outcomes, as indicated by Cappa et al. (2021). "Veracity" refers to the imperative of relying on confident and reliable interpretations of large, diverse, and up-to-date data, a notion echoed by Del Vecchio et al. (2018). In the context of the NPD process, high-veracity data plays a crucial role in reducing the uncertainty that typically characterizes radical innovation efforts. First, it offers validated and trustworthy signals about customer needs that may not yet be explicitly expressed or codified, thereby facilitating the anticipation of latent demands before competitors can act (Urbinati et al., 2018). Second, the use of reliable data minimizes the likelihood of false positives in opportunity identification—a common pitfall in radical innovation, where managers might otherwise be misled by noise or anecdotal evidence (Cooper et al., 2023). Third, high-veracity data supports more confident and evidence-based resource allocation toward projects that are inherently high-risk and potentially high-reward (Cappa et al., 2021). In this sense, it is plausible to hypothesize that the higher the veracity of big data, the higher the intensity of radical innovation. Therefore, we posit the following relationships:

H4. Big data veracity has a positive effect on firms' radical innovation.

[2] In addition to the widely acknowledged dimensions of volume, variety, velocity, and veracity, recent research has proposed further dimensions such as value (Cappa et al., 2021) and variability (Ghasemaghaei and Calic, 2019). Data value refers to the extent to which data - often derived from volume, velocity, and variety - can generate meaningful insights for decision-making, while data variability captures the inconsistency and unpredictability in data flows, formats, semantics, and quality. Despite the growing interest in these additional dimensions, our study focuses on the core Vs—volume, velocity, variety, and veracity—as they remain the most consistently adopted in empirical research on big data. This choice is also aligned with the specific research gap we found in the literature and highlighted by Mariani et al. (2023, p. 21), who call for further exploration of "what are the features of big data (volume, variety, velocity, veracity, etc.) that are most likely to support incremental and radical innovation."

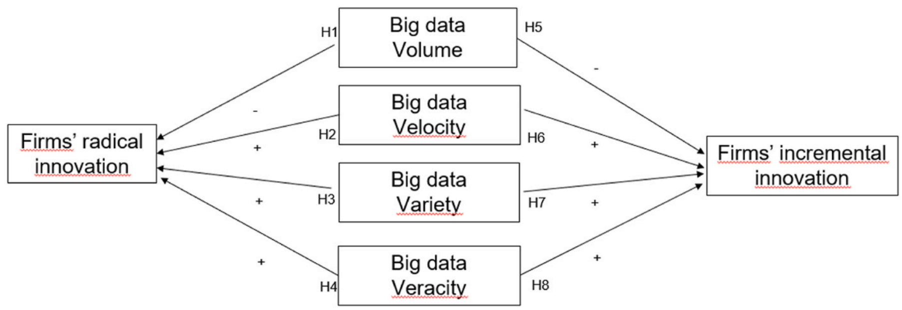

Fig. 1. Research framework.

2.2.2. The "4Vs" of big data and firms' incremental innovation

Moving to firms' incremental innovation, existing studies highlight that the "4Vs" of big data (i.e. volume, variety, velocity, and veracity) can represent a source of knowledge for introducing minor changes to existing products in the market (Mikalef et al., 2019; Cappa et al., 2021). As regards big data volume, existing studies show that the exponential increase in big data volume can negatively affect firms' incremental innovation processes (Ghasemaghaei and Calic, 2020). While large datasets promise enhanced decision-making capabilities, their sheer volume can overwhelm firms, leading to inefficiencies in extracting actionable insights (Ghasemaghaei and Calic, 2020). Incremental innovation depends on targeted, context-specific information, while excessive volume may obscure relevant signals within redundant or irrelevant information (Mikalef et al., 2019). Moreover, recent studies highlight the operational bottlenecks caused by data overload (Acciarini et al. 2023). High volumes of data require substantial computational and human resources, diverting focus from refining existing processes or products (Sivarajah et al., 2024). Additionally, biases in dataset construction may lead firms to prioritize misleading insights, resulting in inefficient innovation strategies (Chang, 2021). In this regard, the cognitive limitations of decision-makers further exacerbate these challenges. Employees tasked with processing large volumes of information may face difficulties identifying relevant patterns, delaying or derailing iterative improvements (Prescott, 2016). Moreover, managing extensive datasets often necessitates costly infrastructure upgrades, which may divert funds from incremental innovation activities (Tan and Zhan, 2017). Therefore, in line with the above arguments, we hypothesize the following relationship:

H5. Big data volume has a negative effect on firms' incremental innovation.

The variety of big data, encompassing diverse sources, such as customer feedbacks, market trends, and internal operations, plays an important role in driving incremental innovation. Firms can integrate heterogeneous data—such as customer feedback, operational metrics, and market signals—to identify subtle trends and opportunities for gradual refinement of products or processes (Mikalef et al., 2019). This diversity enhances adaptive learning and enables more nuanced, fine-tuned innovations that respond to evolving customer needs, which is essential for sustaining incremental innovation (Ghasemaghaei and Calic, 2020). Scholars suggest that the breadth of data variety enables firms to identify incremental opportunities that might not be apparent from a single data source, facilitating refined decision-making and agility in responding to consumer preferences (Capurro et al., 2021). Moreover, firms can leverage diverse data sources to increase their capacity for adaptive learning, allowing them to incrementally evolve their offerings in alignment with changing market conditions (Mikalef et al., 2021). Therefore, we hypothesize that:

H6. Big data variety has a positive effect on firms' incremental innovation.

Big data velocity, characterized by the speed at which data is processed and analyzed, positively influences firms' incremental innovation by enabling timely decision-making and quick adaptation to market changes (Ferrigno et al., 2024; Mikalef et al., 2019). The rapid processing of real-time data allows firms to quickly identify gaps in their operations or products, which in turn facilitates continuous improvements (Ghasemaghaei and Calic, 2019). Rapid data flows also foster dynamic knowledge exchange, both internally and externally, reinforcing the agility needed for incremental innovation in competitive contexts (Erevelles et al., 2016; Dagnino et al., 2021). By accelerating decision-making cycles, big data velocity fosters quicker responses, enhances adaptability, and drives ongoing, small-scale improvements critical for firms' incremental innovation. Consequently, we surmise that:

H7. Big data velocity has a positive effect on firms' incremental innovation.

Big data veracity, referring to the accuracy and trustworthiness of data, is equally vital driving firms' incremental innovation (Acciarini et al., 2023; Cappa et al., 2021; Wamba et al., 2021). Accurate and reliable data allows firms to make informed decisions, reduce uncertainty, and minimize the risk of costly errors (Cappa et al., 2021). When data is trustworthy, firms can make more confident refinements to products or internal processes, minimizing the risk of missteps and reinforcing the effectiveness of innovation routines (Acciarini et al., 2023). Moreover, Veracity also enhances trust in inter-organizational collaboration, enabling coordinated incremental innovations across the value chain (Cappa et al., 2021). Lastly, high trustworthy data enables firms to identify improvement opportunities with greater precision, supporting the continuous, small-scale innovations that characterize incremental innovation (Wamba et al., 2021). As firms increasingly rely on accurate data, they can enhance their decision-making processes and continuously refine their existing offerings, thereby sustaining incremental innovation. Therefore, we hypothesize that:

H8. Big data veracity has a positive effect on firms' incremental innovation.

3. Data and methodology

3.1. Data collection

To test our hypotheses, we collected data on the "4Vs" of big data from Italian firms. In particular, we decided to consider the Italian context because in this country a significant percentage of firms are transitioning toward Industry 4.0 (I4.0) (Messeni Petruzzelli et al., 2022; Mise, 2020). Furthermore, during the last 10 years, the Italian Government has launched several plans and ad-hoc interventions with the aim of stimulating the digitalization of firms and the related adoption of I4.0 technologies, which include big data (Ferrigno et al., 2024; Martinelli et al., 2021).

We then developed our survey on the basis of the existing literature (see Appendix) and before launching it we decided to conduct three preliminary tests to enhance the quality and the clarity of the questionnaire, thus increasing the reliability of the data collected (Crick et al., 2023). In March 2023 we conducted a dual pre-testing phase, involving both scholars in the fields of NPD and big data (five scholars), as well as decision-makers and firms' founders that operate in several sectors (five individuals). This pre-test was particularly useful to obtain insights for refining the questions of the surveys, leading to the incorporations of established definitions of radical and incremental innovation (Mikalef et al., 2019), as well as introducing more sophisticated and clear questions toward big data dimensions. Moreover, in April 2023, a field service company (that was not involved in the administration of the survey) reviewed the questionnaire. Then, the survey was validated through multiple rounds of testing with five companies that utilize big data and with five academics specialized in big data. In June 2023 we started the data collection by relying on the support of an external professional agency which targeted a randomly selected subset of 801 firms. The subset of firms selected by the external professional agency represented both SMEs and large firms, operating in different sectors, and representing the geographical distribution of companies at national level. Between June and August 2023, we contacted the founders and CEOs of these firms using the Computer-Assisted Telephone Interview (CATI) method and we asked them to fill the survey. Out of these 801, 210 firms declared to adopt big data and 155 of them completed the questionnaire, 135 of which were SMEs (less than 250 employees) and 20 were large firms.

To test whether the final sample size was sufficient, we calculated the minimum number of necessary samples for a finite population which is the estimated number of Italian firms. The equation for calculating the sample size is computed as follows:

$$\textit{Minimum sample size} = \frac{N}{1 + \frac{z^2 \times \widehat{p}(1-\widehat{p})}{\varepsilon^2 N}}$$

where N represents the population size, z is the z score for the confidence interval, ε is the margin of error, and p the population proportion which represents the percentage of the value associated with the survey. In our case, N is the estimated number of Italian firms which is approximately 6 million, the confidence interval is set at 99 %, while the margin of error is set at 10 %. We infer the population proportion of Italian firms implementing big data by calculating the ratio of sampled firms' implementing big data over the total number of firms contacted. Then, we set the population proportion at 26.22 % (210/801), which provides 129 firms as a minimum sample size for our survey.

3.2. Constructs and measures

3.2.1. Dependent variables

In line with Subramaniam and Youndt (2005) and Mikalef et al. (2019), we measure radical innovation (*RI*) and incremental innovation (*II*) through a three-items, 7-point Likert scale. Specifically, we asked the following question "Would your firm be able to generate the following innovations?" and proposed the following items to address *RI*:

RAD1: "Innovations which make our prevailing product lines obsolete"
RAD2: "Innovations which radically change our prevailing product lines"
RAD3: "Innovations which make our experience in the prevailing product lines obsolete" and the following items for *II*:
INC1: "Innovations that strengthen our core product/service lines"
INC2: "Innovations that strengthen our existing expertise in prevailing products/services"
INC3: "Innovations that strengthen the way we currently compete"

3.2.2. Independent variables

To measure the "4Vs" of big data, we rely on previous studies on big data characteristics (Ghasemaghaei, 2019, 2021; Ghasemaghaei and Calic, 2020). In line with Johnson et al. (2017) and Ghasemaghaei and Calic (2019, 2020), we measure *volume*, *velocity* and *variety* according to a 7-point Likert scale (ranging from 1, "strongly disagree", and 7, "strongly agree"). On the other hand, to operationalize *veracity*, we rely on the work of Ghasemaghaei (2021). The Likert scale permits us to assess the magnitude of data collected (*volume*) (Wamba et al., 2015; Cappa et al., 2021), the diversity of data per observation (*variety*) (Wamba et al., 2015; Cappa et al., 2021), the speed of collecting and analyzing data (*velocity*) (Ghasemaghaei and Calic, 2020), and the reliability of data collection and analysis (*veracity*) (Ghasemaghaei, 2021). The measures of *volume* and *velocity* consist of four items, whereas those of *variety* and *veracity* consist of three items. Specifically, we asked the following items for *volume*:

VOL1: "In my firm, we analyze large amounts of data"
VOL2: "In my firm, we explore large amounts of data"
VOL3: "In my firm, we use large amounts of data"
VOL4: "In my firm, we explore copious amounts of data"

We asked the following items for *velocity*:

VEL1: "In my firm, we analyze data as soon we receive it"
VEL2: "In my firm, the time period between receiving new data and analysing it is short"
VEL3: "In my firm, we are fast in exploring our data"
VEL4: "In my firm, we analyze data quickly"

Then, we asked the following items for *variety*:

VAR1: "In my firm, we use different data sources to derive useful information"
VAR2: "In my firm, we analyze many type of data"
VAR3: "In my firm, we examine data from a multitude of sources"

Finally, we asked the following items for *veracity*:

VER1: "In my firm, we deal with reliable and precise data"
VER2: "In my company, we analyze high quality data"
VER3: "In my company, we process reliable and consistent data"

3.2.3. Item reduction and validation

We performed exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to extract two factors for the dependent variables (RI and II) and four factors for the independent variables (*volume*, *velocity*, *variety* and *veracity*). Before applying factor analysis, we centered the distribution of each item. The factor loadings are presented in Table 1, which also reports the Cronbach's α, composite reliability (CR) and average variance extracted (AVE) values for all variables.

The factor loadings for all individual items exceed 0.7, which demonstrates strong convergent validity of the scales used in our analysis. Cronbach's α and CR for all variables exceed the threshold of 0.7 (Bagozzi and Youjae, 1988; Black et al., 2010; Garver and Mentzer, 1999), and AVE values exceed 0.5 (Fornell and Larcker, 1981) demonstrating strong convergent validity of the scales used in the analysis and confirming that the scales demonstrate high reliability and internal consistency. Moreover, we performed Harman's single-factor test to assess common method biases by computing AVE for a factor that includes all items as proposed by Podsakoff et al. (2003). AVE is equal to 0.35, which is below the critical threshold of 40 % (Harman, 1976); hence this result indicates that there is no evidence of common method bias.

Table 1
Factor analysis.

Variable/items	Item loadings
***RI* (Cronbach's α = 0.85, CR = 0.865, AVE = 0.543)**	
RAD1	0.72
RAD2	0.82
RAD3	0.89
***II* (Cronbach's α = 0.91, CR = 0.760, AVE = 0.583)**	
INC1	0.84
INC2	0.90
INC3	0.88
***volume* (Cronbach's α = 0.94, CR = 0.832, AVE = 0.584)**	
VOL1	0.91
VOL2	0.87
VOL3	0.89
VOL4	0.89
***velocity* (Cronbach's α = 0.89, CR = 0.732, AVE = 0.576)**	
VEL1	0.71
VEL2	0.77
VEL3	0.92
VEL4	0.92
***variety* (Cronbach's α = 0–87, CR = 0.703, AVE = 0.557)**	
VAR1	0.86
VAR2	0.75
VAR3	0.90
***veracity* (Cronbach's α = 0.87, CR = 0.701, AVE = 0.579)**	
VER1	0.74
VER2	0.95
VER3	0.83

3.2.4. Control variables

We also consider several control variables to control firm specific characteristics. We introduce the centered values of *age*, *revenues* and *size* (proxied by the number of employees) to control for firms' balance sheets. To address the role of human capital, we introduce a categorical variable to control for the *percentage of STEM employees* subdivided in 6 classes: 1) under 5 %, 2) between 5 and 10 %, 3) between 11 and 20 % 4) between 21 and 30 %, 5) between 31 and 50 %, and 6) more than 50 %. In the model, we also account for the *percentage of R&D expenditures* to control for the resources devoted to the accumulation of knowledge, by introducing five classes: 1) under 5 %, 2) between 5 and 10 %, 3) between 11 and 20 %, 4) between 21 and 30 %, and 5) more than 30 %. To control for the presence of previous technological knowledge, we introduce the number of patents present in the *patent portfolio* as a categorical variable subdivided into four classes: 1) no patents; 2) one patent, 3) between two and four patents; 4) more than four patents. Finally, we computed a dummy variable, *High tech sector*, which takes the value 1 when the firm operates in a high-tech sector, and 0 otherwise.

Tables 2 and 3 report the descriptive statistics and the frequency distribution of the categorical variables, respectively.

Table 4 reports GVIF and correlation matrix for the dependent and the independent variables.

Table 2
Descriptive statistics of continuous variables.

	Obs	Mean	SD	Min	Max
II	155	0	0.95	−2.49	1.32
RI	155	0	0.93	−1.70	2.33
volume	155	0	0.97	−2.72	1.35
variety	155	0	0.94	−2.34	1.43
velocity	155	0	0.96	−2.47	1.77
veracity	155	0	0.96	−2.21	1.49
Age	155	0	1.00	−1.66	3.37
Revenues	155	0	1.00	−0.30	9.14
Size	155	0	1.00	−0.25	10.73

Table 3
Frequency distribution of categorical variables.

Percentage of STEM employees:			
- below 5 %	- betw. 5 % and 10 %	- betw. 11 % and 20 %	- betw. 21 % and 30 %
77	26	17	14
- betw. 31 % and 50 %	- more than 50 %		
12	9		
Percentage of R&D expenditures:			
- below 5 %	- betw. 5 % and 10 %	- betw. 11 % and 20 %	- betw. 21 % and 30 %
81	51	15	4
- more than 30 %			
4			
Patent portfolio:			
- no patent	−1 patent	- betw. 2 and 4	- more than 4
112	7	15	21
High tech sector:			
- Low tech sector	- High tech sector		
128	27		
Dummy SME:			
- SMEs	- Large firms		
135	20		

3.3. Methodology

We perform the econometric analysis by estimating an ordinary least squares (OLS) regression as follows:

$$Y_i = Big\ data\ Vs_i \times Dummy\ SME_i + Controls_i + \epsilon_i$$

where Y_i is *RI* or *II*, *Big data* Vs_i represents the big data characteristics (i.e.: *volume, velocity*, *variety* and *veracity*), while *Dummy* SME_i is an interaction term which splits the sample between SMEs and large firms. Then, $Controls_i$ collects all the control variables (i.e.: *age*, *revenues, size, percentage of STEM employees*, *percentage of R&D expenditures*, *patent portfolio,* and *high tech sector*).

4. Results

Table 5 reports the results for the OLS estimation of *RI*. Column 1 includes the estimation results of control variables and it shows that *percentage of STEM employees between 5 % and 10 %*, percentage *of R&D expenditure between 5 % and 10 %,* and *patent portfolio with more than four patents* have a positive and statistically significant effect on *RI* ($\beta = 0.5705$ p-value<0.01, $\beta = 0.3068$ p-value<0.1, and, $\beta = 0.3976$ p-value<0.1, respectively). From column 2 to column 5 the results for each big data Vs (i.e.: *volume*, *velocity*, *variety* and *veracity*) are reported; they have been estimated separately and we find that *velocity*, *variety* and *veracity* positive and statistically significant for SMEs ($\beta = 0.2723$ p-value<0.01, $\beta = 0.1862$ p-value<0.05, and, $\beta = 0.2686$ p-value<0.01, respectively). Then, Column 6 shows the estimation results of the model including Vs altogether. We find that *variety* and *veracity* have a positive and statistically effect on *RI* for SMEs ($\beta = 0.1853$ p-value<0.1, and $\beta = 0.1885$ p-value<0.05, respectively), thus confirming H2 and H4. On the other hand, we could not support H1 and H3 since *volume* has a negative but non-statistically significant effect on *RI* for SMEs, while *velocity* has a positive but non-statistically significant effect on *RI* for SMEs.

Table 6 reports the results for the OLS estimation of *II*. Column 1 includes the estimation results of control variables and it shows that every category of *patent portfolio* has a positive and statistically significant effect on *II* ($\beta = 0.8353$ p-value<0.01, $\beta = 0.5491$ p-value<0.05, and $\beta = 0.7042$ p-value<0.01, respectively). Then, from column 2 to column 5 are reported the results for each big data Vs (i.e.: *volume*, *velocity*, *variety* and *veracity*) estimated separately and we find that each V is positive and statistically significant for SMEs ($\beta = 0.3332$ p-value<0.01, $\beta = 0.4020$ p-value<0.01, $\beta = 0.2743$ p-value<0.01, and $\beta =$

Table 4
GVIF and correlation matrix.

	GVIF	*II*	*RI*	*volume*	*velocity*	*variety*	*veracity*	*Age*	*Revenues*	*Size*
II		1								
RI		0.4	1							
volume	1.451361	0.35	0.24	1						
velocity	1.344698	0.29	0.2	0.5	1					
variety	1.37671	0.42	0.31	0.6	0.32	1				
veracity	1.227135	0.47	0.3	0.37	0.47	0.35	1			
Age	1.11342	0	−0.13	0.02	−0.09	0.03	−0.09	1		
Revenues	2.997798	−0.01	−0.08	0.01	−0.12	0.06	−0.12	0.27	1	
Size	2.709019	0	−0.07	−0.04	−0.13	0.04	−0.12	0.22	0.92	1
Percentage of STEM employees:	1.084048									
- betw. 5 % and 10 %		0.1	0.25	0.13	0.06	0.13	0.05	−0.11	−0.07	−0.05
- betw. 11 % and 20 %		0	0	0.1	0.1	0.03	−0.01	0.05	0.26	0.21
- betw. 21 % and 30 %		−0.02	0.02	−0.04	−0.02	0.06	0.01	−0.01	−0.07	−0.04
- betw. 31 % and 50 %		0.15	−0.08	−0.1	0.1	−0.01	0.13	−0.08	−0.07	−0.03
- more than 50 %		−0.08	−0.03	0.06	−0.07	0.13	−0.02	−0.01	0.11	0.13
Percentage of R&D expenditures:	1.072567									
- betw. 5 % and 10 %		−0.02	0.2	0	0.04	0.05	−0.06	−0.03	−0.13	−0.08
- betw. 11 % and 20 %		0.08	−0.06	0.02	0.01	−0.05	0.04	−0.14	−0.07	−0.05
- betw. 21 % and 30 %		0.08	0.03	0.05	0.11	−0.04	0.07	−0.1	−0.02	−0.01
- more than 30 %		0.03	0.03	0.02	0.2	0.06	0.19	−0.07	−0.04	−0.03
Patent portfolio:	1.093143									
−1 patent		0.18	0	0.01	0.03	0.08	0.02	0.05	−0.05	−0.03
- betw. 2 and 4		0.12	0.08	−0.09	0.02	0.04	−0.03	0.03	0.04	−0.01
- more than 4		0.19	0.12	0.1	−0.09	0.2	0.04	0.23	0.1	0.15
High tech sector	1.182093	0.12	−0.01	−0.04	−0.02	−0.08	0.14	−0.04	−0.1	−0.04
Dummy SME	1.281352	0.02	0.02	0.01	0.11	0.05	0.07	−0.12	−0.52	−0.4

	Percentage of STEM employees:					Percentage of R&D expenditures:			
	- betw. 5 % and 10 %	- betw. 11 % and 20 %	- betw. 21 % and 30 %	- betw. 31 % and 50 %	- more than 50 %	- betw. 5 % and 10 %	- betw. 11 % and 20 %	- betw. 21 % and 30 %	- more than 30 %
Percentage of STEM employees:									
- betw. 5 % and 10 %	1								
- betw. 11 % and 20 %	−0.16	1							
- betw. 21 % and 30 %	−0.14	−0.11	1						
- betw. 31 % and 50 %	−0.13	−0.1	−0.09	1					
- more than 50 %	−0.11	−0.09	−0.08	−0.07	1				
Percentage of R&D expenditures:									
- betw. 5 % and 10 %	0.16	−0.11	0.16	0.05	0	1			
- betw. 11 % and 20 %	−0.03	0.16	−0.1	0.07	−0.08	−0.23	1		
- betw. 21 % and 30 %	0.04	−0.06	−0.05	0.11	−0.04	−0.11	−0.05	1	
- more than 30 %	−0.07	0.07	0.09	−0.05	0.13	−0.11	−0.05	−0.03	1
Patent portfolio:									
−1 patent	−0.01	−0.08	0.04	0.29	−0.05	−0.02	0.03	−0.04	−0.04
- betw. 2 and 4	0.03	0.02	−0.03	0.07	0.11	0.14	−0.11	−0.05	0.08
- more than 4	−0.08	0.1	0.07	−0.04	−0.02	0.04	0.06	0.17	−0.06
High tech sector	0.02	−0.11	0.03	0.31	0.18	0.11	−0.09	0.03	0.14
Dummy SME	0.02	−0.11	−0.01	0.04	0.01	0.02	0.13	−0.06	0.06

	Patent portfolio:				
	−1 patent	- betw. 2 and 4	- more than 4	High tech sector	Dummy SME
Patent portfolio:					
−1 patent	1				
- betw. 2 and 4	−0.07	1			
- more than 4	−0.09	−0.13	1		
High tech sector	0.06	0.08	−0.08	1	
Dummy SME	0.08	−0.13	0.04	−0.13	1

0.4416 p-value<0.01, respectively). However, the estimation of big data characteristics cannot be interpreted separately hence we run an OLS by estimating all Vs. The results are shown in Column 6. On the one hand, we find that *variety* and *veracity* have a positive and statistically significant effect on II for SMEs ($\beta = 0.2018$ p-value<0.05, and $\beta = 0.3347$ p-value<0.01, respectively), thus confirming H6 and H8. On the other hand, we could not support H5 and H7 since *volume* and *velocity* have a positive but non-statistically significant effect on *II* for SMEs.

5. Robustness check

As robustness check, we estimate the linear regression models by

Table 5
RI OLS estimation results.

	Dep. variable: RI					
Ind. variables:	(1)	(2)	(3)	(4)	(5)	(6)
volume						
- large firms		0.4968***				0.4434***
		(0.1679)				(0.135)
- SMEs		0.1724				−0.0370
		(0.1090)				(0.1718)
variety						
- large firms			0.3320*			0.1007
			(0.1683)			(0.1695)
- SMEs			0.2723***			0.1853*
			(0.0851)			(0.0979)
velocity						
- large firms				0.2442		−0.2663
				(0.2275)		(0.1699)
- SMEs				0.1862**		0.0561
				(0.0922)		(0.1233)
veracity						
- large firms					0.5987***	0.5588***
					(0.164)	(0.1769)
- SMEs					0.2686***	0.1885**
					(0.0714)	(0.0856)
Age	−0.0855	−0.0886	−0.0733	−0.0977	−0.0809	−0.1009
	(0.0992)	(0.1015)	(0.0938)	(0.1072)	(0.0979)	(0.1024)
Revenues	0.0169	−0.0300	−0.0226	0.0108	−0.0063	−0.028
	(0.0827)	(0.0828)	(0.0832)	(0.083)	(0.0727)	(0.0742)
Size	−0.0512	−0.0176	−0.043	−0.0235	0.0529	0.0194
	(0.1278)	(0.1136)	(0.1192)	(0.1242)	(0.1223)	(0.1033)
Percentage of STEM employees:						
- betw. 5 % and 10 %	0.5705***	0.4821**	0.4263**	0.5236**	0.4877**	0.4167**
	(0.2107)	(0.1896)	(0.1961)	(0.2072)	(0.2013)	(0.1895)
- betw. 11 % and 20 %	0.1066	0.0510	0.0734	0.0254	0.0839	0.0704
	(0.2167)	(0.2135)	(0.2175)	(0.229)	(0.2092)	(0.2191)
- betw. 21 % and 30 %	0.0274	0.0339	−0.0649	0.0188	0.0205	0.0107
	(0.2547)	(0.2513)	(0.2842)	(0.2381)	(0.2464)	(0.2685)
- betw. 31 % and 50 %	−0.214	−0.1386	−0.3024	−0.315	−0.3457	−0.2737
	(0.3498)	(0.3551)	(0.3419)	(0.3994)	(0.3369)	(0.3429)
- more than 50 %	−0.0368	−0.1060	−0.2351	−0.0079	−0.015	−0.1167
	(0.3439)	(0.3313)	(0.3219)	(0.3339)	(0.3461)	(0.3617)
Percentage of R&D expenditures:						
- betw. 5 % and 10 %	0.3068*	0.2859*	0.2977*	0.2670	0.3212**	0.3041*
	(0.1615)	(0.1467)	(0.1511)	(0.1653)	(0.1572)	(0.1595)
- betw. 11 % and 20 %	−0.0810	−0.1074	−0.0439	−0.0928	−0.1092	−0.0865
	(0.1836)	(0.1896)	(0.1758)	(0.187)	(0.1567)	(0.173)
- betw. 21 % and 30 %	0.1224	−0.0146	0.2308	−0.0576	0.0133	−0.0142
	(0.4226)	(0.496)	(0.4884)	(0.4822)	(0.4945)	(0.5447)
- more than 30 %	0.3636	0.2646	0.1983	0.0997	0.0636	−0.0268
	(0.5934)	(0.6084)	(0.5987)	(0.5698)	(0.5907)	(0.6088)
Patent portfolio:						
1 patent	0.2565	0.1911	0.1190	0.2570	0.2795	0.1617
	(0.3433)	(0.3334)	(0.3397)	(0.3546)	(0.3142)	(0.3146)
- betw. 2 and 4	0.2314	0.2973	0.2105	0.233	0.2748	0.2848
	(0.2992)	(0.2972)	(0.2931)	(0.2844)	(0.2543)	(0.2824)
- more than 4	0.3976*	0.3651	0.2288	0.4606*	0.3454	0.2949
	(0.2315)	(0.2378)	(0.2435)	(0.2418)	(0.2417)	(0.2476)
High tech sector	−0.0428	−0.0459	0.0372	−0.0029	−0.2113	−0.1432
	(0.2768)	(0.2731)	(0.2699)	(0.2773)	(0.2804)	(0.2734)
Constant	0.2180	0.2598	0.3402	0.1832	−0.1513	0.1656
	(0.5266)	(0.5211)	(0.5064)	(0.5617)	(0.5450)	(0.5057)
Observations	155	155	155	155	155	155
R-squared	0.1248	0.1778	0.1921	0.1598	0.2290	0.2710
F statistics	1.44	2.80***	2.48**	1.98**	6.55***	9.58***

Clustered-robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

accounting for sample selection (Heckman, 1976). Firstly, we run a probit model on the 210 firms who declared to adopt big data in which the dependent variable is a dummy that takes the value 1 if the firms completed the questionnaire, and 0 otherwise. Then, we control for the *percentage of STEM employees* and *sector*. Lastly, we estimate OLS regression models on the 155 firms that completed the questionnaire to obtain coefficients robust to sample selection. Table 7 reports the results for the OLS estimation of *RI* and *II* that account for sample bias selection. The results further confirm the support of H3, H4, H6 and H8.

6. Discussion

In this paper, we contribute to the literature on firms' NPD processes by exploring if and how radical and incremental innovations are associated with firms' big data. More precisely, adopting a KBV, we unpacked big data into four dimensions: volume, variety, velocity, and veracity. Then, we examined whether these dimensions affect firms' radical and incremental innovations. With this objective, we conducted an OLS regression on a sample of 155 Italian firms' survey responses.

Table 6
II OLS estimation results.

	Dep. Variable: II					
Ind. variables:	(1)	(2)	(3)	(4)	(5)	(6)
volume						
- large firms		0.6086***				0.2494*
		(0.216)				(0.148)
- SMEs		0.3332***				0.0833
		(0.0927)				(0.1022)
variety						
- large firms			0.6417***			0.3679**
			(0.2045)			(0.1703)
- SMEs			0.4020***			0.2018**
			(0.0888)			(0.0999)
velocity						
- large firms				0.5930**		0.1331
				(0.2729)		(0.2416)
- SMEs				0.2743***		0.0043
				(0.0862)		(0.0919)
veracity						
- large firms					0.5484**	0.3413**
					(0.2118)	(0.1588)
- SMEs					0.4416***	0.3347***
					(0.0796)	(0.1021)
Age	−0.0294	−0.0339	−0.0114	−0.0331	−0.0182	−0.0052
	(0.1104)	(0.1046)	(0.0958)	(0.1072)	(0.0913)	(0.0864)
Revenues	0.0347	−0.0412	−0.0312	0.023	−0.0124	−0.0626
	(0.0664)	(0.0577)	(0.0541)	(0.0612)	(0.066)	(0.0559)
Size	−0.0087	0.0388	0.0103	0.0656	0.1067	0.1165
	(0.0993)	(0.0853)	(0.0811)	(0.0976)	(0.0985)	(0.0885)
Percentage of STEM employees:						
- betw. 5 % and 10 %	0.3042	0.1553	0.0737	0.2185	0.1885	0.0457
	(0.2078)	(0.1734)	(0.1737)	(0.1722)	(0.1581)	(0.1435)
- betw. 11 % and 20 %	−0.0493	−0.1443	−0.1133	−0.1881	−0.0686	−0.1371
	(0.3087)	(0.3066)	(0.3275)	(0.3094)	(0.275)	(0.292)
- betw. 21 % and 30 %	−0.0202	−0.0185	−0.1537	−0.0543	−0.0608	−0.1307
	(0.2514)	(0.259)	(0.281)	(0.2334)	(0.2426)	(0.2721)
- betw. 31 % and 50 %	0.2558	0.3568	0.1092	0.0701	0.0318	0.0214
	(0.2995)	(0.3019)	(0.2682)	(0.3308)	(0.2667)	(0.2519)
- more than 50 %	−0.3748	−0.5031*	−0.6927**	−0.3358	−0.3349	−0.5458
	(0.3404)	(0.2924)	(0.3031)	(0.3092)	(0.3721)	(0.3471)
Percentage of R&D expenditures:						
- betw. 5 % and 10 %	−0.1355	−0.1645	−0.1598	−0.2065	−0.0843	−0.1256
	(0.1841)	(0.1711)	(0.1707)	(0.163)	(0.1617)	(0.1591)
- betw. 11 % and 20 %	0.2091	0.166	0.2612	0.1909	0.175	0.197
	(0.1905)	(0.1783)	(0.204)	(0.2092)	(0.2087)	(0.2182)
- betw. 21 % and 30 %	0.1434	−0.0334	0.3065	−0.1605	0.0431	0.0698
	(0.5582)	(0.4454)	(0.4444)	(0.5815)	(0.4517)	(0.4396)
- more than 30 %	0.2838	0.1216	0.0183	−0.1111	−0.2412	−0.3174
	(0.3093)	(0.385)	(0.3502)	(0.3319)	(0.4288)	(0.4402)
Patent portfolio:						
1 patent	0.8353***	0.7354**	0.6276*	0.833***	0.8665***	0.7183***
	(0.2864)	(0.294)	(0.3341)	(0.2558)	(0.2192)	(0.2719)
- betw. 2 and 4	0.5491**	0.6688***	0.5274***	0.5239***	0.6023***	0.5950***
	(0.2178)	(0.2124)	(0.1954)	(0.1904)	(0.172)	(0.1838)
- more than 4	0.7042***	0.6349***	0.4556**	0.7869***	0.6237***	0.4999**
	(0.1963)	(0.2085)	(0.2248)	(0.2161)	(0.1983)	(0.2321)
High tech sector	0.2946	0.2747	0.4346**	0.3675*	0.1192	0.2635
	(0.2034)	(0.2042)	(0.2019)	(0.2057)	(0.1856)	(0.1796)
Constant	−0.2356	−0.1312	−0.0536	−0.4650	−0.5961	−0.4536
	(0.5447)	(0.5167)	(0.4703)	(0.5689)	(0.5088)	(0.5026)
Observations	155	155	155	155	155	155
R-squared	0.1558	0.2852	0.3135	0.2532	0.3462	0.4257
F statistic	1.94***	4.60***	4.35***	3.59***	4.50***	8.17***

Clustered-robust standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

The first finding of this study is that the variety dimension of big data positively affects both radical and incremental innovation, providing support for H2 and H6. This positive effect implies that firms must leverage a diverse range of data observations and several sources (Ghasemaghaei and Calic, 2020; Huang et al., 2018) to introduce breakthrough innovation, but also to improve already existing products. In other words, managers should recognize the importance of "smart" over "large" volume of data, as articulated by Cappa et al. (2021) and George and Martine (2014), in which the relevance and quality of the source of knowledge become pivotal factors.

Then, in support of H4 and H8, our results show that the veracity dimension, i.e. the reliability and insightfulness of data (Cappa et al., 2021), is linked to increased firms' radical and incremental innovation. This finding is in line with studies that highlight the need to analyze the effects of veracity on firms' innovation (Ghasemaghaei and Calic, 2020). The outcome suggests that firms will develop new knowledge beneficial for their innovation performance when big data and their sources are not biased, and hence their analysis produces accurate results. Our results

Table 7
Robustness check estimation results.

Dep. variables:	RI	II
Ind. variables:	(1)	(2)
Big Data Volume		
- large firms	0.5177***	0.2576*
	(0.1308)	(0.1545)
- SMEs	−0.0551	0.0718
	(0.1639)	(0.0936)
Big Data Velocity		
- large firms	0.0397	0.3879**
	(0.1454)	(0.1790)
- SMEs	0.1768*	0.1621*
	(0.0961)	(0.0906)
Big Dara Variety		
- large firms	−0.2978**	0.1290
	(0.1399)	(0.2298)
- SMEs	0.0534	−0.0159
	(0.103)	(0.0896)
Big Data Veracity		
- large firms	0.5872***	0.3787**
	(0.1574)	(0.158)
- SMEs	0.1627**	0.2761***
	(0.083)	(0.0837)
Age	−0.1148	0.0619
	(0.1018)	(0.1184)
Revenues	0.0150	−0.0810
	(0.0773)	(0.0604)
Size	−0.0554	0.1405
	(0.1291)	(0.0922)
Percentage of STEM employees:		
- betw. 5 % and 10 %	0.5201***	0.2214
	(0.1801)	(0.1592)
- betw. 11 % and 20 %	0.0869	−0.0552
	(0.2596)	(0.2597)
- betw. 21 % and 30 %	−0.0184	−0.1177
	(0.2552)	(0.264)
- betw. 31 % and 50 %	−0.1901	0.0592
	(0.3132)	(0.2541)
- more than 50 %	0.1367	−0.0863
	(0.3553)	(0.3315)
Percentage of R&D expenditures:		
- betw. 5 % and 10 %	0.4030***	−0.0710
	(0.1503)	(0.1379)
- betw. 11 % and 20 %	0.0012	0.2520
	(0.1484)	(0.2399)
- betw. 21 % and 30 %	−0.2628	−0.0174
	(0.3004)	(0.4559)
- more than 30 %	−0.0338	−0.2978
	(0.5698)	(0.4625)
Patent portfolio:		
1 patent	0.1734	10.019***
	(0.2482)	(0.2646)
- betw. 2 and 4	0.2911	0.6749***
	(0.2886)	(0.1616)
- more than 4	0.4340**	0.4787*
	(0.1834)	(0.2645)
High tech sector	−0.2279	0.0514
	(0.2580)	(0.1767)
Constant	0.5559	−0.4176
	(0.4951)	(0.5197)
Sample selection correction	Yes	Yes
Observations	155	155
R-squared	0.2365	0.4438
F statistic	9.12***	9.74***

Clustered-robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

thus highlight that the veracity of big data, through efforts in data analysis, is beneficial for firms' NPD by allowing the identification of reliable insights from big data. The above-mentioned outcomes evidence that big data can be a valuable source of knowledge that positively impacts firms' radical and incremental innovations when it is characterized by variety and veracity.

In summary, this study highlights that, regardless of the type of innovation (radical or incremental) that firms aim to introduce in their market, the variety and veracity of big data play a critical role for enhancing firms' NPD processes. Specifically, leveraging diverse and multifaceted data sources (variety), alongside ensuring the reliability and accuracy of data (veracity), also enables firms to drive both breakthrough innovations and refinements to existing products. These findings emphasize the need for managers to prioritize data diversity and quality in order to maximize innovation outcomes. In the subsections below, we explore the contributions, limitations, and potential avenues for further development inherent in this study.

6.1. Implications for theory

Despite the considerable attention from both academics and practitioners to the effects of big data on NPD (Ghasemaghaei and Calic, 2020; Johnson et al., 2017; Mikalef et al., 2019), we do not know much about the theoretical mechanisms and empirical quantification of their impact on firms' radical and incremental innovation. In fact, some big data scholars noted that a research area that requires investigation is understanding "what are the features of big data (volume, variety, velocity, veracity, etc.) that are most likely to support incremental vs. radical innovation" (Mariani et al., 2023, p. 21) and suggested that "the effects of big data on firm performance could be further analyzed by considering firm-specific factors, such as the types of innovations delivered to the market, for example, architectural versus modular and/or incremental versus radical (Cappa et al., 2021, p. 62). The principal impetus behind our study lays in responding to these calls, resulting in the articulation of significant contributions outlined below.

Therefore, first, we underscored the significance of employing an appropriate theoretical foundation to comprehend the circumstances influencing the beneficial or detrimental impact of big data on firms' radical and incremental innovation. We followed recent calls about big data (Bstieler et al., 2018; Cappa et al., 2021; Mariani et al., 2023) and to this end we have adopted the KBV (Cooper et al., 2023; Grant, 1996). This approach provides a framework to deepen our understanding of what form big data should take for it to be transformed into a distinct, valuable knowledge that impacts firms' radical and incremental innovation. Specifically, we posited that big data could be conceptualized as intangible assets that yield valuable knowledge for firms' NPD processes. This KBV perspective on big data is highly important, given the escalating abundance of data and the increasing levels of knowledge that can arise from analyzing customers' data (Cooper et al., 2023; de Camargo et al., 2018). We hope that both scholars and managers will embrace this view of big data to better understand how to generate knowledge to launch new products that can better satisfy customers' needs.

Second, we undertook an examination of the "multidimensionality of big data" (Cappa et al., 2021; Tambe, 2014), to empirically assess what form it should take to be a valuable knowledge for firms' NPD. This facet of our study advances existing research that takes a holistic approach in examining the impact of big data on firms' radical and incremental innovation (Mikalef et al., 2019). More specifically, while Mikalef et al. (2019) found a correlation between big data analytics capability and radical innovation, they did not explore the specific mechanisms through which big data analytics capability acts as a catalyst for such innovations, leaving room for a more in-depth investigation. Therefore, we contribute to this research by dissecting big data into different dimensions. Building upon the work of Johnson and colleagues (2017), we focused on three established dimensions of big data—volume, variety, and velocity—and additionally expanded our inquiry to include a fourth dimension, veracity, in line with recent scholarly works (Cappa et al., 2021; Ghasemaghaei and Calic, 2020). We contended that the volume of big data negatively impacts radical and incremental innovation, while the variety, velocity and veracity of big data have a positive effect. Using an OLS regression on a sample of 155 Italian firms, our findings revealed that both variety and veracity positively affect firms' both radical and incremental innovation while volume and velocity do not exert any statistically significant effect. These results enable us to contribute to the work of Mariani et al. (2023). In fact, these authors underscore the need for deeper inquiry into understanding what are the features of big data (volume, variety, velocity, veracity, etc.) that are most likely to support incremental and radical innovation (Mariani et al., 2023: p. 21). In our study, we found that variety and veracity support both radical and incremental innovation.

Third, the findings of our study provide compelling insights into the application of paradox theory (Smith and Lewis, 2011) for understanding how the distinct characteristics of big data influence firms' innovation strategies. Specifically, our results reveal that certain big data attributes simultaneously affect both radical and incremental innovation, suggesting that big data serves as a paradox enabler within the innovation spectrum. This observation aligns with the broader discourse in innovation management, which identifies the trade-off between radical and incremental innovation as a central paradox (March 1991; Tushman and O'Reilly, 1996). In a nutshell, our findings suggest that the paradox theory framework, which emphasizes the coexistence of seemingly contradictory yet interdependent elements, might be a robust lens for unpacking this phenomenon (Smith and Lewis, 2011). In the context of big data, its ability to foster both exploration (radical innovation) and exploitation (incremental innovation) underscores its paradoxical role in innovation processes. For instance, the data-driven insights enabling incremental improvements can simultaneously reveal disruptive opportunities for radical transformation. This duality challenges firms to navigate tensions effectively, balancing the short-term gains of exploitation with the long-term potential of exploration (Benner and Tushman, 2003; Gupta et al., 2006).

6.2. Implications for practice

Our study also offers three important managerial implications, specifically devoted to firms' decision-makers. First, the results from our empirical analysis indicate fruitful suggestions for the improvement of firms' innovation performance. More specifically, our findings suggest firms' decision-makers to specifically focus on big data technologies' implementation according to their variety, and veracity dimensions. Conversely, focusing on the improvement of the amount of big data per se might constitute an inconsequential activity with negligible impact on innovation performance. Therefore, firms' decision-makers should channel organizational investments towards "smart" rather than merely "big" data (George and Martine, 2014). In particular, regarding the development of new products, "smart" data can stimulate a more efficient usage of the scare resources that characterize the small business context.

Second, our findings evidence that the potential adoption of "smart" data can help firms to introduce innovative products on the market, thereby advancing their positioning within the industry. Considering these findings, firms' decision-makers should devote considerable attention to the strategic development of big data technologies, particularly within the framework of a specific NPD strategy. This strategic focus can enable them to capture substantial market shares, fostering both the spread of their radical and incremental innovations and their individual innovation and financial performance.

Third, the findings of our study suggest that firms' decision-makers should actively involve customers' (and in general, stakeholders') in generating new data and the simultaneous hiring/retaining of HR resources, in particular employees with specific skills in big data analytics. This fact can, on one side, improve the quantity (i.e., "big" data) and the quality of data (i.e., "smart" data) within the firm and, on the other side, enhance the effectiveness of the big data analytics within the firm.

6.3. Limitations and future research directions

Our research is not free from limitations, some of which can be interpreted as possible research opportunities. First, we unpack the influence of big data on firms' NPD by examining the impact of big data volume, variety, velocity, and veracity on firms' radical and incremental innovation. By doing so, we contribute to previous research calling for a more comprehensive "understanding of big data's effects by considering other dimensions of big data" (Cappa et al., 2021, p. 63). Despite this contribution, we recognize that other dimensions can be studied, such as data value and data variability. For instance, value—understood as the ability to derive economic benefit from data—has increasingly been recognized as a relevant feature of Big Data (Ghasemaghaei and Calic, 2019). Nevertheless, a number of studies contend that value should not be treated as a characteristic per se, but rather as an outcome or endogenous variable resulting from effective Big Data use (when considered together with the other Vs). Data variability—defined as the inconsistency and unpredictability of data in terms of flows, formats, semantics, and quality (Katal et al., 2013)—represents another key yet underexplored characteristic of Big Data. Although its operationalization remains methodologically challenging, data variability may gain increasing relevance, as it exposes firms to diverse and evolving data inputs that may reveal unmet customer needs or novel use cases, thus fostering the introduction of breakthrough or revised products in the market. Future studies could explore these dimensions of big data to further understand the effects of big data on firms' NPD.

Second, we have analyzed the impact of big data on firms' radical and incremental innovation. Albeit we provide a comprehensive understanding of whether big data characteristics support radical and incremental innovation (Mariani et al., 2023), we are aware that "the effects of big data on firm performance could be further analyzed by considering firm-specific factors, such as the types of innovations delivered to the market, for example, architectural versus modular" (Cappa et al., 2021, p. 62; based on the work of Cammarano et al., 2019). Moreover, prior research has highlighted that firms' innovation performance can be shaped by their open innovation search strategies (Del Sarto et al., 2023). Concurrently, other studies have demonstrated that specific characteristics of big data can affect how firms engage in open innovation (Ferrigno et al., 2024). Building on these insights, we propose that it would be worthwhile to investigate whether certain big data characteristics—such as value and veracity—moderate the relationship between open innovation search strategies and firms' radical and incremental innovation performance.

Third, we have explored the impacts of big data on NPD within the Italian context. We acknowledge that the generalizability of our findings is confined to this particular type of organization operating in the specified country. Subsequent studies could investigate the effects of big data dimensions on firms' innovations across different countries. Similarly, we recognize that most of the firms we analyzed are SMEs. Future research might investigate how big data dimensions influence radical and incremental innovations in other types of firms, such as large enterprises, or start-ups.

Fourth, in investigating these dimensions within this specific context, we utilized surveys and self-reported measures, a method widely established in previous big data research (Johnson et al., 2017). However, we acknowledge the necessity for observable and measured indices (Cappa et al., 2021). Given our awareness of the limitations inherent in survey-based approaches, we intentionally structured our sample survey to gather responses from firms' decision-makers. Nevertheless, it is important to recognize that our findings rely on the perceptions of individuals, albeit ones who are knowledgeable informants. As a result, we advocate for future big data research that incorporates more observable and measurable data to enhance transparency and replicability.

Finally, we observe the presence of numerous effects. This suggests the possibility of an interactive effect among the big data dimensions influencing the outcome, in our case radical and incremental innovation. Future studies could benefit from employing QCA to investigate how the combination of multiple big data dimensions impacts radical and/or incremental innovation performance. QCA offers a more in-depth exploration of complex relationships between variables, providing the opportunity to identify patterns and interactions that might escape more traditional analytical methods (Ragin, 2008). This approach could yield a more comprehensive and detailed understanding of how various big data dimensions collectively influence firms' radical and incremental innovation performance.

CRediT authorship contribution statement

Giulio Ferrigno: Writing – review & editing, Writing – original draft, Project administration, Investigation, Conceptualization. **Saverio Barabuffi:** Writing – original draft, Methodology, Formal analysis. **Enrico Marcazzan:** Writing – original draft, Data curation. **Andrea Piccaluga:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the participants of the SIMA Conference 2024 in Parma, the R&D Management Conference 2024 in Stockholm and the research seminar at Scuola Superiore Sant'Anna for their constructive comments on earlier versions of this paper. We also acknowledge funding from the EU Horizon Europe research and innovation programme under Grant Agreement n. 101079043, SoBigData RI PPP. All errors are our responsibility.

APPENDIX A. Operationalization of the variables

Variable	Definition	Operationalization	Database	Methodological reference
Radical innovation	Major departures from existing capabilities in the firm, and constitute the basis for completely new products, services or business models (Mikalef et al. 2019)	7-point Likert scale: RAD1: "Innovations which make our prevailing product lines obsolete" RAD2: "Innovations which radically change our prevailing product lines" RAD3: "Innovations which make our experience in the prevailing product lines obsolete"	Survey firms' decision-makers	Mikalef et al. (2019)
Incremental innovation	Minor departures from existing capabilities in the firm, and constitute the basis for completely new products, services or business models (Mikalef et al. 2019)	7-point Likert scale: INC1: "Innovations that strengthen our core product/service lines" INC2: "Innovations that strengthen our existing expertise in prevailing products/services" INC3: "Innovations that strengthen the way we currently compete"	Survey firms' decision-makers	Mikalef et al. (2019)
Volume	Amount of data collected (Wamba et al., 2015; Cappa et al., 2021)	7-point Likert scale: (1) in my firm, we analyze large amounts of data; (2) in my firm, the quantity of data we explore is substantial; (3) in my firm, we use a great deal of data; (4) in my firm, we scrutinize copious volumes of data	Survey firms' decision-makers	Ghasemaghaei and Calic (2019); Ghasemaghaei and Calic (2020); Ghasemaghaei (2021); Johnson et al. (2017)
Variety	Assortment of data collected per observation (Cappa et al., 2021; Wamba et al., 2015)	7-point Likert scale: (1) in my firm, we use several different sources of data to gain insights; (2) in my firm, we analyze many types of data; (3) in my firm, we examine data from a multitude of sources	Survey firms' decision-makers	Ghasemaghaei and Calic (2019); Ghasemaghaei and Calic (2020); Ghasemaghaei (2021); Johnson et al. (2017)
Velocity	The speed and the frequency of processing and integrating data (Ghasemaghaei and Calic, 2020)	7-point Likert scale: (1) in my firm, we analyze data as soon as we receive it; (2) in my firm, the time period between when we get new data and when we analyze it is short; (3) in my firm, we are fast in exploring our data; (4) in my firm, we analyze data speedily	Survey firms' decision-makers	Ghasemaghaei and Calic (2019); Ghasemaghaei and Calic (2020); Johnson et al. (2017); Ghasemaghaei (2021)
Veracity	The reliability and insightfulness of data (Cappa et al., 2021)	7-point Likert scale: (1) Within my firm, one has to deal with certain and precise data; (2) Within my firm, high quality data is analyzed; (3) Within my firm, reliable and consistent data is processed in my firm, we analyze data as soon as we receive it	Survey firms' decision-makers	Cappa et al. (2021); Ghasemaghaei and Calic (2019)
R&D Exp	Percentage of the firm's R&D expenses on total revenues	0 (percentage lower than 5 %); 1 (between 5 % and 10 %); 2 (between 11 % and 20 %); 3 (between 21 % and 30 %); 4 (higher than 30 %)	Survey firms' decision-makers	Cassetta et al. (2020); Messeni Petruzzelli et al. (2022) ; Zahra and George (2002)
Age	Firm's age at the time of the survey	1–284	Survey firms' decision-makers	Kelly and Amburgey (1991) Kelly and Amburgey (1991); Messeni-Petruzzelli et al. (2022)
Size	Firm's size at the time of the survey	1–250	Survey firms' decision-makers	Arbore and Ordanini (2006); Horváth and Szabó (2019) ; Messeni Petruzzelli et al. (2022)
Patent Portfolio	Presence of patents in firm's patent portfolio	1) no patents; 2) 1 patent, 3) between 2 and 4 patents; 4) more than 4 patents	Survey firm' decision-makers	Martinelli et al. (2021); Damij et al. (2022)

(*continued on next page*)

(*continued*)

Variable	Definition	Operationalization	Database	Methodological reference
Tech int ind	Technological intensity of the firm's industry	1 (manufacturing SMEs with a NACE code classified as high-technology or medium high-technology, or service SMEs with a NACE code classified as knowledge intensive services); 0 otherwise	Survey firms' decision-makers	Messeni-Petruzzelli et al. (2022)

Data availability

Data will be made available on request.

References

Acciarini, Chiara, Cappa, Francesco, Boccardelli, Paolo, Raffaele Oriani, R., 2023. How can organizations leverage big data to innovate their business models? A systematic literature review. Technovation 123, 1–18. https://doi.org/10.1016/j.technovation.2023.102713.

Ahmad, Sohel, Mallick, Debasish N., Schroeder, Roger G., 2013. New product development: impact of project characteristics and development on performance. J. Prod. Innovat. Manag. 30, 331–348. https://doi.org/10.1111/j.1540-5885.2012.01002.x.

Andriopoulos, Constantine, Lewis, Marianne W., 2009. Exploitation-exploration tensions and organizational ambidexterity: managing paradoxes of innovation. Organ. Sci. 20 (4), 696–717. https://doi.org/10.1287/orsc.1080.0406.

Arbore, A., Ordanini, A., 2006. Broadband divide among SMEs: The role of size, location and outsourcing strategies. International Small Business Journal 24 (1), 83–99.

Atuahene-Gima, Kwaku, 2005. Resolving the capability–rigidity paradox in new product innovation. J. Market. 69 (4), 61–83. https://doi.org/10.1509/jmkg.2005.69.4.61.

Bagozzi, Richard P., Youjae, Yi, 1988. On the evaluation of structural equation models. J. Acad. Market. Sci. 16, 74–94. https://doi.org/10.1007/BF02723327.

Benner, Mary J., Tushman, Michael L., 2003. Exploitation, exploration, and process management: the productivity dilemma revisited. Acad. Manag. Rev. 28 (2), 238–256. https://doi.org/10.5465/amr.2003.9416096.

Bharadwaj, Neeraj, Noble, Charles H., 2017. Finding innovation in data rich environments. J. Prod. Innovat. Manag. 34 (5), 560–564. https://doi.org/10.1111/jpim.12407.

Black, William C., Babin, Barry J., Anderson, Rolph E., 2010. Multivariate Data Analysis: A Global Perspective. Pearson.

Bstieler, Ludwig, Gruen, Thomas, Akdeniz, Billur, Brick, Danielle, Du, Shuili, Guo, Lin, Khanlari, Moein, McIllroy, James, O'Hern, Matt, Yalcinkaya, Goksel, 2018. Emerging research themes in innovation and new product development: insights from the 2017 PDMA-UNH doctoral consortium. J. Prod. Innovat. Manag. 35 (3), 300–307. https://doi.org/10.1111/jpim.12447.

Cammarano, Antonello, Michelino, Francesca, Caputo, Mauro, 2019. Open innovation practices for knowledge acquisition and their effects on innovation output. Technol. Anal. Strat. Manag. 31 (11), 1297–1313. https://doi.org/10.1080/09537325.2019.1606420.

Cappa, Francesco, Oriani, Raffaele, Peruffo, Enzo, McCarthy, Ian, 2021. Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance. J. Prod. Innovat. Manag. 38 (1), 49–67. https://doi.org/10.1111/jpim.12545.

Capurro, Rosita, Fiorentino, Raffaele, Garzella, Stefano, Giudici, Alessandro, 2021. Big data analytics in innovation processes: which forms of dynamic capabilities should be developed and how to embrace digitization? Eur. J. Innovat. Manag. 25 (6), 273–294. https://doi.org/10.1108/EJIM-05-2021-0256.

Cassetta, E., Monarca, U., Dileo, I., Di Berardino, C., Pini, M., 2020. The relationship between digital technologies and internationalisation. Evidence from Italian SMEs. Industry and Innovation 27 (4), 311–339.

Chang, Victor, 2021. An ethical framework for big data and smart cities. Technol. Forecast. Soc. Change 165, 1–11. https://doi.org/10.1016/j.techfore.2020.120559.

Chen, Hsinchun, Chiang, Roger H.L., Storey, Veda C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36 (4), 1165–1188. https://doi.org/10.2307/41703503.

Cooper, Robert G., 1983. A process model for industrial new product development. IEEE Trans. Eng. Manag. 30 (1), 2–11. https://doi.org/10.1109/TEM.1983.6448637.

Cooper, Robert G., 1990. Stage-gate systems: a new tool for managing new products. Bus. Horiz. 33 (3), 44–54. https://doi.org/10.1016/0007-6813(90)90040-I.

Cooper, Robert G., 2018. The drivers of success in new-product development. Ind. Mark. Manag. 76, 36–47. https://doi.org/10.1016/j.indmarman.2018.07.005.

Cooper, R.G., Kleinschmidt, E.J., 1986. An investigation into the new product process: steps, deficiencies, and impact. J. Prod. Innovat. Manag. 3 (2), 71–85.

Cooper, C., Pereira, V., Vrontis, D., Liu, Y., 2023. Extending the resource and knowledge based view: insights from new contexts of analysis. J. Bus. Res. 156, 113523.

Crick, James M., Crick, Dave, Ferrigno, Giulio, 2023. Coopetition and the marketing/entrepreneurship interface in an international arena. Int. J. Entrepreneurial Behav. Res. https://doi.org/10.1108/IJEBR-01-2022-0099 (in press).

Dagnino, Giovanni Battista, Picone, Pasquale Massimo, Giulio, Ferrigno, 2021. Temporary competitive advantage: a state-of-the-art literature review and research directions. Int. J. Manag. Rev. 23 (1), 85–115. https://doi.org/10.1111/ijmr.12242.

Davenport, Thomas H., Stephan, Kudyba, 2016. Designing and developing analytics-based data products. MIT Sloan Manag. Rev. 58 (1), 83–88.

Davenport, Thomas H., Paul, Barth, Bean, Randy, 2012. How "big data" is different. MIT Sloan Manag. Rev. 54 (1), 43–46.

de Camargo, Fiorini, Paula, Roman Pais Seles, Bruno Michel, Chiappetta Jabbour, Charbel Jose, Barberio Mariano, Enzo, de Sousa Jabbour, Ana Beatriz Lopes, 2018. Management theory and big data literature: from a review to a research agenda. Int. J. Inf. Manag. 43, 112–129. https://doi.org/10.1016/j.ijinfomgt.2018.07.005.

Dean, Tereza, Zhang, Haisu, Xiao, Yazhen, 2023. Use big data to leverage customer need diversity for radical innovation. J. Market. Manag. 39 (15–16), 1620–1644. https://doi.org/10.1080/0267257X.2023.2273277.

Del Vecchio, Pasquale, Di Minin, Alberto, Messeni Petruzzelli, Antonio, Panniello, Umberto, Pirri, Salvatore, 2018. Big data for open innovation in SMEs and large corporations: trends, opportunities, and challenges. Creativ. Innovat. Manag. 27 (1), 6–22. https://doi.org/10.1111/caim.12224.

Dobbs, Richard, Manyika, James, Woetzel, Jonathan, 2015. The Four Global Forces Breaking All the Trends. McKinsey & Company.

Du, Rex Yuxing, Kamakura, Wagner A., 2012. Quantitative trendspotting. J. Mark. Res. 49 (4), 514–536. https://doi.org/10.1509/jmr.10.0167.

Dwivedi, Yogesh K., Krishna, Venkitachalam, Sharif, Amir M., Al-Karaghouli, Wafi, Vishanth, Weerakkody, 2011. Research trends in knowledge management: analyzing the past and predicting the future. Inf. Syst. Manag. 28 (1), 43–56. https://doi.org/10.1080/10580530.2011.536112.

Eling, Katrin, Griffin, Abbie, Langerak, Fred, 2016. Consistency matters in formally selecting incremental and radical new product ideas for advancement. J. Prod. Innovat. Manag. 33, 20–33. https://doi.org/10.1111/jpim.12320.

Erevelles, Sunil, Fukawa, Nobuyuki, Swayne, Linda, 2016. Big Data consumer analytics and the transformation of marketing. J. Bus. Res. 69 (2), 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001.

Ferrigno, Giulio, Barabuffi, Saverio, Marcazzan, Enrico, Piccaluga, Andrea, 2024. What "V" of the big data influence SMEs' open innovation breadth and depth? An empirical analysis. R. Manag. https://doi.org/10.1111/radm.12727.

Fornell, Claes, Larcker, David F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. 18 (1), 39–50. https://doi.org/10.1177/002224378101800104.

Garver, Michael S., Mentzer, John T., 1999. Logistics research methods: employing structural equation modeling to test for construct validity. J. Bus. Logist. 20 (1), 33–57.

George, Gerard, Martine, R., 2014. Haas, and alex pentland. "Big data and management.". Acad. Manag. J. 57 (2), 321–326. https://doi.org/10.5465/amj.2014.4002.

Ghasemaghaei, Maryam, 2019. Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. Enterp. Inf. Syst. 13 (5), 650–674. https://doi.org/10.1080/17517575.2019.1576228.

Ghasemaghaei, Maryam, 2021. Understanding the impact of big data on firm performance: the necessity of conceptually differentiating among big data characteristics. Int. J. Inf. Manag. 57, 102055. https://doi.org/10.1016/j.ijinfomgt.2019.102055.

Ghasemaghaei, Maryam, Calic, Goran, 2019. Does big data enhance firm innovation competency? The mediating role of data-driven insights. J. Bus. Res. 104, 69–84. https://doi.org/10.1016/j.jbusres.2019.07.006.

Ghasemaghaei, Maryam, Calic, Goran, 2020. Assessing the impact of big data on firm innovation performance: big data is not always better data. J. Bus. Res. 108, 147–162. https://doi.org/10.1016/j.jbusres.2019.09.062.

Grant, Robert M., 1996. Toward a knowledge-based theory of the firm. Strateg. Manag. J. 17 (S2), 109–122. https://doi.org/10.1002/smj.4250171110.

Gupta, Manjul, George, Joey F., 2016. Toward the development of a big data analytics capability. Inf. Manag. 53 (8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004.

Gupta, Anil K., Smith, Ken G., Shalley, Christina E., 2006. The interplay between exploration and exploitation. Acad. Manag. J. 49 (4), 693–706. https://doi.org/10.5465/amj.2006.22083026.

Harman, Harry H., 1976. Modern Factor Analysis. University of Chicago press.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann. Econ. Soc. Meas. 5 (number 4), 475–492. NBER.

Horváth, D., Szabó, R.Z., 2019. Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities? Technological forecasting and social change 146, 119–132.

Huang, Xueli, Soutar, Geoffrey N., Brown, Alan, 2002. New product development processes in small and medium-sized enterprises: some Australian evidence. J. Small Bus. Manag. 40 (1), 27–42. https://doi.org/10.1111/1540-627X.00036.

Huang, Cheng-Kui, Wang, Tawei, Huang, Tzu-Yen, 2018. Initial evidence on the im-pact of big data implementation on firm performance. Inf. Syst. Front. 22, 1–13. https://doi.org/10.1007/s10796-018-9872-5.

Jagtap, Sandeep, Nguyen Khanh Duong, Linh, 2019. Improving the new product development using big data: a case study of a food company. Br. Food J. 121 (11), 2835–2848. https://doi.org/10.1108/BFJ-02-2019-0097.

Johnson, Jeff S., Friend, Scott B., Lee, Hannah S., 2017. Big data facilitation, utilization, and monetization: exploring the 3Vs in a new product development process. J. Prod. Innovat. Manag. 34 (5), 640–658. https://doi.org/10.1111/jpim.12397.

Katal, Avita, Wazid, Mohammad, Goudar, Rayan H., 2013. Big data: issues, challenges, tools and good practices. 2013 Sixth International Conference on Contemporary Computing (IC3). IEEE, pp. 404–409. https://doi.org/10.1109/IC3.2013.6612229.

Kelly, D., Amburgey, T.L., 1991. Organizational inertia and momentum: A dynamic model of strategic change. Academy of management journal 34 (3), 591–612.

March, James G., 1991. Exploration and exploitation in organizational learning. Organ. Sci. 2 (1), 71–87. https://doi.org/10.1287/orsc.2.1.71.

Mariani, Marcello M., Machado, Isa, Magrelli, Vittoria, Dwivedi, Yogesh K., 2023. Artificial intelligence in innovation research: a systematic review, conceptual framework, and future research directions. Technovation 122 (1), 102623. https://doi.org/10.1016/j.technovation.2022.102623.

Martinelli, Arianna, Mina, Andrea, Moggi, Massimo, 2021. The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. Ind. Corp. Change 30 (1), 161–188. https://doi.org/10.1093/icc/dtaa060.

Martinelli, A., Mina, A., Moggi, M., 2021. The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. Industrial and Corporate Change 30 (1), 161–188.

Marzi, Giacomo, Ciampi, Francesco, Dalli, Daniele, Dabic, Marina, 2021. New product development during the last ten years: the ongoing debate and future avenues. IEEE Trans. Eng. Manag. 68 (1), 330–344. https://doi.org/10.1109/TEM.2020.2997386.

Matthias, Olga, Fouweather, Ian, Gregory, Ian, Vernon, Andy, 2015. Making sense of Big Data – can it transform operations management? Int. J. Oper. Prod. Manag. 37 (1), 37–55. https://doi.org/10.1108/IJOPM-02-2015-0084.

McAfee, Andrew, Brynjolfsson, Erik, 2012. Big data: the management revolution. Harv. Bus. Rev. 90 (10), 60–68.

Messeni Petruzzelli, A., Murgia, G., & Parmentola, A. 2022. How can open innovation support SMEs in the adoption of I4. 0 technologies? An empirical analysis. *R&D Management*, *52*(4), 615-632.

Mikalef, Patrick, Pappas, Ilias O., Krogstie, John, Giannakos, Michail, 2018. Big data analytics capabilities: a systematic literature review and research agenda. Inf. Syst. E Bus. Manag. 16, 547–578. https://doi.org/10.1007/s10257-017-0362-y.

Mikalef, Patrick, Boura, Maria, Lekakos, George, Krogstie, John, 2019. Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. Br. J. Manag. 30, 272–298. https://doi.org/10.1111/1467-8551.12343.

Mikalef, Patrick, van de Wetering, Rogier, Krogstie, John, 2021. Building dynamic capabilities by leveraging big data analytics: the role of organizational inertia. Inf. Manag. 58 (6), 103412.

Mise, 2020. "Piano Transizione 4.0 (2019-2020)". mimit.gov. Available at: https://www.mise.gov.it/images/stories/documenti/Slide_transizione_22giugno2020.pdf.

Nafizah, Ully Y., Roper, Stephen, Mole, Kevin, 2024. Estimating the innovation benefits of first-mover and second-mover strategies when micro-businesses adopt artificial intelligence and machine learning. Small Bus. Econ. 62, 411–434. https://doi.org/10.1007/s11187-023-00779-x.

Petruzzelli, Messeni, Antonio, Murgia, Gianluca, Parmentola, Adele, 2022. How can open innovation support SMEs in the adoption of I4.0 technologies? An empirical analysis. R. Manag. 52 (4), 615–632. https://doi.org/10.1111/radm.12507.

Podsakoff, Philip M., MackEnzie, Scott B., Lee, Jeong-Yeon, Podsakoff, Nathan P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88 (5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Prescott, Michael E., 2016. Big data: innovation and competitive advantage in an information media analytics company. J. Innov. Manag. 4 (1), 92–113. https://doi.org/10.24840/2183-0606_004.001_0007.

Radicic, Dragana, Petković, Saša, 2023. Impact of digitalization on technological innovations in small and medium-sized enterprises (SMEs). Technol. Forecast. Soc. Change 191, 122474. https://doi.org/10.1016/j.techfore.2023.122474.

Ragin, Charles C., 2008. Measurement versus calibration: a set-theoretic approach. In: The Oxford Handbook of Political Methodology. Oxford Academic, pp. 174–198. https://doi.org/10.1093/oxfordhb/9780199286546.003.0008.

Rakshit, Sandip, Mondal, Sandeep, Islam, Nazrul, Jasimuddin, Sajjad, Zhang, Zuopeng, 2021. Social media and the new product development during COVID-19: an integrated model for SMEs. Technol. Forecast. Soc. Change 170, 120869. https://doi.org/10.1016/j.techfore.2021.120869.

Ritala, Paavo, Hurmelinna-Laukkanen, Pia, 2013. Incremental and radical innovation in coopetition—the role of absorptive capacity and appropriability. J. Prod. Innovat. Manag. 30 (1), 154–169. https://doi.org/10.1111/j.1540-5885.2012.00956.x.

Ritala, Paavo, Keränen, Joona, Fishburn, Jessica, Ruokonen, Mika, 2024. Selling and monetizing data in B2B markets: four data-driven value propositions. Technovation 130, 102935. https://doi.org/10.1016/j.technovation.2023.102935.

Salomo, Sören, Talke, Katrin, Strecker, Nanja, 2008. Innovation field orientation and its effect on innovativeness and firm performance. J. Prod. Innovat. Manag. 25 (6), 560–576. https://doi.org/10.1111/j.1540-5885.2008.00322.x.

Sivarajah, U., Kumar, S., Kumar, V., Chatterjee, S., Li, J., 2024. A study on big data analytics and innovation: from technological and business cycle perspectives. Technol. Forecast. Soc. Change 202, 123328.

Smith, Wendy K., Lewis, Marianne W., 2011. Toward a theory of paradox: a dynamic equilibrium model of organizing. Acad. Manag. Rev. 36 (2), 381–403. https://doi.org/10.5465/amr.2009.0223.

Subramaniam, Mohan, Youndt, Mark A., 2005. The influence of intellectual capital on the types of innovative capabilities. Acad. Manag. J. 48, 450–463. https://doi.org/10.5465/amj.2005.17407911.

Tambe, Prasanna, 2014. Big data investment, skills, and firm value. Manag. Sci. 60 (6), 1452–1469. https://doi.org/10.1287/mnsc.2014.1899.

Tan, Kim Hua, Zhan, Yuanzhu, 2017. Improving new product development using big data: a case study of an electronics company. R. Manag. 47 (4), 570–582.

Tushman, Michael L., O'Reilly III, Charles A., 1996. Ambidextrous organizations: managing evolutionary and revolutionary change. Calif. Manag. Rev. 38 (4), 8–30. https://doi.org/10.2307/41165852.

Urbinati, Andrea, Bogers, Marcel, Chiesa, Vittorio, Frattini, Federico, 2018. Creating and capturing value from Big Data: a multiple-case study analysis of provider companies. Technovation 84, 21–36. https://doi.org/10.1111/radm.12242.

Wamba, Samuel Fosso, Akter, Shahriar, Edwards, Andrew, Chopin, Geoffrey, Gnanzou, Denis, 2015. How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. 165, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031.

Wamba, Samuel Fosso, Epie Bawack, Ransome, Cameron, Guthrie, Queiroz, Maciel M., Kevin, Daniel André Carillo, 2021. Are we preparing for a good AI society? A bibliometric review and research agenda. Technol. Forecast. Soc. Change 164, 120482. https://doi.org/10.1016/j.techfore.2020.120482.

Zahra, S.A., George, G., 2002. Absorptive capacity: A review, reconceptualization, and extension. Academy of management review 27 (2), 185–203.

Zhang, Haisu, Wu, Fang, Cui, Anna Shaojie, 2015. Balancing market exploration and market exploitation in product innovation: a contingency perspective. Int. J. Res. Market. 32 (3), 297–308. https://doi.org/10.1016/j.ijresmar.2015.03.004.