Big data, machine learning and uncertainty in foresight studies

Vinicius Muraro and Sergio Salles-Filho

Vinicius Muraro is based at the Department of Business Administration, Lund University, Lund, Sweden. Sergio Salles-Filho is based at the Department of Science and Technology Policy, Universidade Estadual de Campinas, Campinas, Brazil.

Abstract

Purpose – *Currently, foresight studies have been adapted to incorporate new techniques based on big data and machine learning (BDML), which has led to new approaches and conceptual changes regarding uncertainty and how to prospect future. The purpose of this study is to explore the effects of BDML on foresight practice and on conceptual changes in uncertainty.*

Design/methodology/approach – *The methodology is twofold: a bibliometric analysis of BDML-supported foresight studies collected from Scopus up to 2021 and a survey analysis with 479 foresight experts to gather opinions and expectations from academics and practitioners related to BDML in foresight studies. These approaches provide a comprehensive understanding of the current landscape and future paths of BDML-supported foresight research, using quantitative analysis of literature and qualitative input from experts in the field, and discuss potential theoretical changes related to uncertainty.*

Findings – *It is still incipient but increasing the number of prospective studies that use BDML techniques, which are often integrated into traditional foresight methodologies. Although it is expected that BDML will boost data analysis, there are concerns regarding possible biased results. Data literacy will be required from the foresight team to leverage the potential and mitigate risks. The article also discusses the extent to which BDML is expected to affect uncertainty, both theoretically and in foresight practice.*

Originality/value – *This study contributes to the conceptual debate on decision-making under uncertainty and raises public understanding on the opportunities and challenges of using BDML for foresight and decision-making.*

Keywords *Uncertainty, Futures studies, Foresight, Big data, Machine learning, Artificial intelligence*

Paper type *Research paper*

1. Introduction

Across decades, futures studies have adapted approaches and tools to cope with uncertain and complex contexts at sectoral, regional and national scales. Over the past 40 years, the term "foresight" has gained widespread usage, denoting studies facilitating evidence-based policies and strategies, both at governmental and corporate levels. Foresight seeks a collective future vision among stakeholders, using diverse qualitative and quantitative methodologies, including artificial intelligence tools such as big data and machine learning (BDML) (Miles *et al.*, 2008, 2016).

BDML has been changing the way that many industries and sectors think about the future (Xu *et al.*, 2018). Big data refers to massive data sets and their tools to support acquiring, managing and processing information. Machine learning algorithms, including large language models (LLM) such as ChatGPT and other chatbots, provide a refined analytical mechanism that looks for structure, classification and hidden patterns in data sets, based on historical (big) data (von der Gracht *et al.*, 2015). However, limitations, such as data quality, data reliability, data interpretability and ethical issues, could impact foresight activities and results (Reimsbach-kounatze, 2015).

There are few studies about the effect on foresight approaches and decision-making in complex environments by using BDML. Some questions remain when discussing the role of BDML for foresight. Given that BDML has important effects on the availability and processing of large amounts of data, how will it affect capabilities and methodological approaches for foresight? Would BDML be able to reshape the concept of uncertainty and the practice of managing uncertainty in decision-making processes?

The objective of this study is to explore the effects of BDML in foresight practice, focusing on understanding how these techniques are changing conceptual and methodological approaches for foresight. The structure of this study is as follows: Section 2 introduces the theoretical background, followed by Section 3 outlining the research design and methods. In Section 4, the results are presented, followed by a discussion in Section 5. The concluding remarks are provided in Section 6.

2. Conceptual background

2.1 Uncertainty and the future

The Second World War highlighted the value of good planning, strategizing and management of complex situations, which led to the formalization of futures studies as an important decision support tool for government and business sectors (Georghiou *et al.*, 2008; Miles, 2010). Because we do not have a historical series of figures about the future, we intuitively base our prospective on the information we have (past and present), try to unfold it and imagine what might happen. Uncertainty is undoubtedly the underlying condition in all kinds of methods and approaches.

There is a vast literature on uncertainty and how it has been considered and managed for prospecting – even predicting – the future and making decisions (Dosi and Egidi, 1991; Knight, 1921; Marchau *et al.*, 2019; Shackle, 1969, 2010). In this study, we adopt a combination of well-known concepts of uncertainty to contribute to one of the objectives presented above: to what extent does BDML affect the perception and nature of uncertainty itself in foresight activities.

The first concept is Frank Knight's classic definition of uncertainty, which is the inevitable condition of the partial knowledge that underlies all decisions, *"neither total ignorance nor complete and perfect information, but partial knowledge"* (Knight, 1921). A second concept, which complements Knight's definition, is that of George Shackle, for whom a decision is always made under bounded uncertainty, meaning "*neither perfect foresight nor chaos*" (Shackle, 1969).

Partial knowledge refers to situations in which information remains incomplete regardless of the contextual factors or available means of inquiry. The information cannot be obtained simply because it does not exist in the present, only in the future. Knight's proposal is a seminal one, as it attempts to distinguish uncertainty from risk. In his proposition, uncertainty refers to the impossibility to know all possible outcomes of a certain event in advance. On the contrary, when all possibilities can be known beforehand, it is not uncertainty, but rather risk – or ambiguity, as argued by several authors later on (Dequech, 2000; Marchau *et al.*, 2019). Risk is calculable, because the variables are fully known *ex ante* and subject to an objective probabilistic distribution (e.g. a roulette wheel). Ambiguity, on the other hand, is the situation where information about the variables is hidden and does not vary over time (a classic example is a box with *n* black balls and *m* red balls) (Dequech, 2000; Ellsberg, 1961).

Under the condition of uncertainty, variables change over time and cannot be known *ex ante*, either because new variables may emerge and change the system, and the methods of calculation and interpretation may also be altered. This is a characteristic of open systems, as pointed out by Prigogine and Stengers (1984). Uncertainty, as defined here, is then a

condition of open systems (variables and possible outputs vary over time). Risk and ambiguity are conditions of closed systems (variables and possible outputs do not vary over time). Human and social affairs are, by definition, open systems (Metcalfe *et al.*, 2021; Prigogine and Stengers, 1984; Shackle, 1969).

Partial knowledge enables us to define an "event horizon," which frames a bounded uncertainty at a given moment in a given context for a given set of eyes. This means that the conditions of uncertainty will be, in practice, bounded by the state of art (available data) and by the imagination (unavailable data) of those involved in the process of framing what is relevant to take into consideration when looking at the future. George Shackle's concept of bounded uncertainty complements Frank Knight's concept of partial knowledge by suggesting that when considering the future, agents form potential scenarios by using known information and envisioning "unknowns" to define the range of possible or subjectively probable outcomes. It remains uncertain in the same sense as Knight's, and it is bounded by the perspectives of the agents. Futures must be envisioned, and decisions be made, that is why "unknowns" must be framed and bounded [1].

Marchau *et al.* (2019) propose a slightly different way of defining uncertainty, ambiguity and risk by introducing "levels of uncertainty." The authors suggest classifying uncertain conditions into $4+1$ levels between total determinism and complete ignorance (similar to Shackle's bounded uncertainty). Level 1 corresponds to the context of "a clear enough future," where only a few deterministic outputs can be predicted (corresponding, in our perspective, to the concept of risk); Level 2 occurs in situations with identifiable alternative futures, treatable by stochastic systems (in our view, similar to ambiguity); Level 3 would arise in contexts with some plausible, non-deterministic futures; Level 4 is divided into two sub-levels: 4a many plausible futures, and 4b unknown futures. Levels 3 and 4 fit in the definition of uncertainty we use here (Metcalfe *et al.*, 2021). This practical way to split uncertainty may help understanding the impacts of BDML over futures studies as we will explore further later in this article.

2.2 Foresight to manage uncertainty

Foresight is defined as a systematic and structured process of thinking, imagining and creating assumptions about the future, exploring trends and potential scenarios that could emerge from a varied source of data and opinions. Different from forecast analysis, which looks for predictions, foresight proposes to reach possible views of the future because it considers that social systems are open, influenced by known and unknown variables. Therefore, foresight has become a popular tool for science, technology and innovation precisely because it proposes a logic for managing uncertainty and building a common perception of the future instead of trying to predict it (Irvine and Martin, 1984; Loveridge, 2009; Miles *et al.*, 2017).

Foresight always operates under conditions of partial knowledge and bounded uncertainty. By definition, being partial and bounded means being context- and time-dependent. The parallel with the concept of event horizon seems to be helpful. An event horizon is the visible limit of an event, beyond which nothing can be seen or detected, except if you trespass it and look back. The only possibility to really see the future is going there. Even with time travel, the future, much like a black hole, remains an unidirectional path; once reached, returning is not possible [2].

This is how human beings deal with social phenomena, which are open and complex systems whose variables, interactions and outcomes cannot be fully known *ex ante*. Resultingly, in foresight studies, agents are always working over data and their interpretation. If we embrace this perspective, in theoretical terms, the application of BDML to futures studies might – very likely – reshape agents' perceptions of how events will unfold, consequently influencing their decisions, not necessarily improving accuracy. The advent of BDML and artificial intelligence introduces the prospect of augmenting the existing informational landscape through an

enormous volume of data, bolstering computational capacity and generating uncountable scenarios of potential outcomes. These new capabilities pose the potential to question the essence of uncertainty itself – pondering whether uncertainty might become risk or ambiguity – but also to recalibrate the methodologies used in futures studies.

An example of a continuous foresight process is presented in Figure 1 and includes five basic steps: pre-foresight, recruitment, generation, action and renewal (Miles, 2002; Popper, 2008) On pre-foresight, the main activities are defining rationales, objectives, project team and the prospective methodologies. Recruiting people (i.e. facilitators, experts and other stakeholders) and collecting data for generating future insights are part of the recruitment phase. Generation is considered the heart of the foresight process, which focuses on the prospective effort of the exercise. In the action phase, the foresight process is up to its primary objective, informing decision-making. Finally, the renewal phase consists of monitoring and evaluating processes to verify if foresight has achieved its goals and prepare foresight's new cycle. Several foresight methods can be included in this process, ranging from more qualitative (e.g. brainstorming, expert panels, workshops and focus groups) to more quantitative approaches (e.g. bibliometrics, trend extrapolation and statistical and economic models). In most of cases, foresight uses multi-method approaches, combining opinion-driven with data-driven techniques and tools to develop informed insight about what the future might hold.

2.3 New data-driven prospective approaches

Data-driven approaches using BDML are receiving increasing attention in literature (Kayser and Blind, 2017; Trappey *et al.*, 2019; Zhou *et al.*, 2020). Big data is defined as a set of data whose size becomes a problem, and the usual collection, storage, management and analysis tools do not fit correctly. It is also often characterized by the 5Vs – volume, variety, velocity, value and veracity (Loukides, 2010). Precisely, machine learning algorithms analyze historical (big) data by acting as adaptative systems that could perform tasks more efficiently through experience. They can continuously learn the patterns from historical data and can make projections about the future. Several methods are currently used in machine learning, such as linear regression, logistic regression, decision trees, support vector machines (SVM), artificial neural networks (ANN) and deep learning (Hastie *et al.*, 2009; Lecun *et al.*, 2015; Shalev-Shwartz and Ben-David, 2014).

Various sources provide valuable data for foresight activities, such as scientific publications, patents, news, social media, websites and other structured and unstructured databases. Regarding algorithms, text mining techniques like natural language processing (NLP) and LLM are used for text analysis, enabling pattern recognition, concept summarization, relationship identification and document classification (Kehl *et al.*, 2020), possibly providing conceptual maps, technological trends and offering insight of social behavior (Amanatidou *et al.*, 2012; Geurts *et al.*, 2022; Glassey, 2012; Pang, 2010). As an

Figure 1 Continuous foresight process
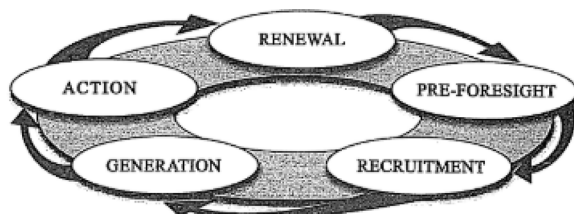
Source: Courtesy of Miles (2002)

example, Spaniol and Rowland (2023) demonstrated that ChatGPT could generate preliminary scenarios, reducing costs and fostering new insights in scenario planning. This approach enhances participants' "futures literacy" and encourages discussing diverse future scenarios. Additionally, Albert et al. (2015) proposed a method to gauge technology maturity through sentiment analysis of blog texts, offering a unique viewpoint compared to patents or scientific literature, as it reflects public opinions. This approach involves collecting blog data, applying sentiment analysis to identify terms related to different technology maturity levels and integrating expert surveys for validation. The literature suggests that foresight methods are reshaped to introduce massive data to support prospective outputs and decision-making. This is made either by adopting it as a slot in consolidated methodologies or by combining methods that use massive data in their processes (Yufei et al., 2016).

To leverage the potential of BDML, Geurts et al. (2022) propose a hybrid AI-expert approach, a reevaluation of the role of experts in foresight, involving a dynamic collaboration between the collective knowledge of human experts and the data-driven capabilities of BDML. They highlight several reasons for this approach: first, as foresight often deals with emerging phenomena, relying only on historical data can confine thinking to past patterns, impeding the identification of novel possibilities (open system, new variables); second, this integration reinforces the validity of findings by encouraging the exploration of new alternative futures and fostering diverse mental models for prospective scenarios; third, the dynamic interplay between human experts and AI mechanisms aids in the identification and rectification of biased outcomes arising from data, by ensuring the contextualization of results within a comprehensive framework. The integration of AI has also the potential to automate aspects of information gathering, analysis and scenario development, fostering a semi-automated continuous foresight process that liberates human capacity for intuition, creativity and imagination (Brandtner and Mates, 2021; Rožanec et al., 2023; Spaniol and Rowland, 2023). As a new generation of foresight studies, Saritas et al. (2022) define "Foresight-on-site," suggesting that it should be carried out closer to the sites of its application, in a continuous and wide participatory model. In this sense, BDML and AI automation in foresight could enhance not only policy formulation but also the operational decisions and actions in response to real-time challenges.

Implementing BDML in foresight studies, while promising, is not without some challenges. Issues such as poor data quality – due to issues of relevance, inaccuracy or inconsistency – coupled with the potential for data manipulation during processes like dimensionality reduction or data cleaning, can bias analyses and distort foresight outcomes (Hagen et al., 2019; L'Heureux et al., 2017). Furthermore, ethical considerations around data privacy and ownership rights compound these challenges (Stahl and Wright, 2018). To effectively navigate these complexities, a comprehensive ethical framework is essential, promoting a responsible application of BDML in foresight initiatives (Mittelstadt and Floridi, 2016).

Besides some limitations, the potential of BDML to impact prospective and futures studies is not negligible. As argued in this study, BDML changes the methodological toolbox of futures studies; in doing so, it may also have effects on the outputs and outcomes of any prospective exercise and consequently, the perceptions about the future.

3. Research design and methods

The methodological approach to explore the effects of the use of BDML tools in foresight is twofold: bibliometric analysis (Hess, 1997), to obtain a current overview of foresight studies supported by BDML and deepen in newly developed methodological approaches, and survey analysis, to raise the perceptions and opinions from foresight experts regarding BDML. Both approaches have been largely used to analyze trends and approach experts' opinions (Cabral et al., 2019; Karaca and Öner, 2015; Keller and von der Gracht, 2014), and combined, they can provide adequate knowledge to discuss current and future

methodological changes and the role of uncertainty in foresight. Detailed methodological steps are presented in Table 1.

3.1 Bibliometrics – current overview of foresight and big data and machine learning

To capture the current state of foresight studies integrated with BDML, we constructed three complementary data sets. The first, "Foresight in Science, Technology, and Innovation" (FSTI database), encompasses papers focused on foresight studies in these domains. The second, "Big Data and Machine Learning" (BDML database), includes publications on the advancements and applications of those technologies. The third, "Foresight in STI with Big Data and Machine Learning" [3] (FSTI+BDML), represents an intersection of the two aforementioned data sets, comprising articles and reviews that embody the combination between foresight studies and BDML technologies, all sourced from Scopus up to the year 2021. Scopus, an Elsevier product, indexes more than 75,000 items from more than 5,000 publishers and 16,000 authors profiles, comprehending a substantial part of academic research in the field. We used the software Vantage Point [4] for uploading, deduplicating and cleaning the data – tasks that included removing incomplete records, correcting errors and standardizing data formats to ensure a robust analysis foundation. NLP techniques, including tokenization (the process of breaking down text or a sequence of characters into smaller units called tokens) and n-grams (sequential combinations of tokens), were used to classify the papers based on their methodological approaches by analyzing the co-occurrence of terms in the abstract related to various foresight methodologies (Popper, 2008) and BDML techniques. This analysis pinpointed the relationships between various foresight methodologies and BDML techniques. To augment the accuracy of our classification, manual reading of selected papers supplemented the NLP-driven process. Furthermore, we compiled a list of corresponding authors along with their email addresses from the FSTI database, which served to conduct a survey analysis, engaging foresight practitioners and experts directly in our study. Tables and charts were created in Microsoft Excel, and Gephi supported network analysis and visualization. Due to the lack of a consolidated foresight studies database, we limited to analyze scientific publications on the topic.

3.2 Experts survey – trends of foresight and big data and machine learning

3.2.1 Projections about effects of big data and machine learning on foresight process. Based on the literature review and inspired by Keller and von der Gracht (2014), we formulated 12 future projections about the implications of BDML in 2025. The set of projections in Table 2 was categorized based on the five steps of foresight process to explore the impact of BDML in different stages of foresight development.

Table 1		Methodological approach: bibliometrics and survey	
	#	Step	Tool/method
Bibliometrics (phase 1)	1	Definition of keywords for searching documents regarding foresight supported by BDML	Literature review
	2	Searching and collecting peer review documents	Scopus
	3	Data cleaning, remove inconsistent and duplicate records	Vantage Point
	4	Classification of methodological approaches by NLP of abstract fields and manual reading	Vantage Point
	5	Data visualization: creation of tables, charts and network graphs	MS Excel/Gephi
Survey (phase 2)	6	Development of projections regarding the impact of BDML in foresight	Literature review
	7	Online survey design	Survey Monkey
	8	Defining target respondents: experts' e-mail address collection	Vantage Point/Python
	9	Send, follow and monitor survey answering	Survey Monkey
	10	Analyze results and data visualization	Vantage Point/Excel
Source: Authors' own creation			

Table 2	Description of 12 projections about influence of BDML		
#	*Foresight step*	*Short title*	*Projection*
1	Pre-foresight	Support objectives definition	Future-oriented activities' objectives will be easily defined using BDML tools for its development in 2025
2	Pre-foresight	Guide methodological choice	The possibility of using BDML tools will make the methodological choice for futures studies easier in 2025
3	Recruitment	Require analytical skills	Data scientists or coding, analyzing and data visualization skills will be required to develop consistent futures studies in 2025
4	Recruitment	Support data collection	The relevant data for future-oriented activities will be less time-consuming and easily accessed using big data tools in 2025
5	Recruitment	Improve data quality	The quality of future-relevant data will be significantly enhanced by the application of BDML tools or techniques in foresight studies in 2025
6	Generation	Enhance data analysis	The quality of future-relevant data analysis will be significantly enhanced by the adoption of BDML tools and techniques in foresight studies in 2025
7	Generation	Drive qualitative methods	BDML tools and techniques will be critical to support qualitative analysis in foresight activity in 2025
8	Generation	Increase data manipulation	The use of big BDML solutions for futures studies will increase the frequency of manipulated (or biased) data in 2025
9	Action	Support results transfer	BDML tools and techniques will increase futures studies' embeddedness to strategic planning or decision-making in 2025
10	Action	Increase decision accuracy	BDML tools and techniques will increase the accuracy of decision-making based on foresight in 2025
11	Renewal	Support foresight evaluation	Evaluating and monitoring future-oriented activities will be easily reached when using BDML tools in 2025
12	Renewal	Support objectives' reach	Futures studies' objectives will be easily reached when BDML tools are used for its development in 2025

Source: Authors' own creation

The initiation of the foresight process, particularly the definition of objectives and methodological choices, is increasingly influenced by access to advanced databases and data analysis tools like BDML. The first two projections (P1, P2), termed pre-foresight projections, address this initial influence. BDML techniques are expected to significantly affect the recruitment step, demanding analytical skills from the team to acquire future-relevant data, assure data quality and identify experts that could contribute to the foresight exercise, leading to projections P3, P4 and P5. Projections related to the generation step (P6, P7, P8) were based on the notion that high-quality data and improved analysis would enhance prospective activities, fostering imagination and complementing qualitative methods. Consequently, various BDML techniques could support the action step (P9, P10) by facilitating the transfer of prospective outputs to decision-making processes and enhancing decision quality. The renewal projections (P11, P12) aim to assess the impact of BDML in monitoring implementation and evaluating the achievement of foresight objectives. These projections underwent internal validation by experts and foresight practitioners.

Given the exponential growth in digital technologies, a short time horizon was chosen to capture both the immediate and near-future impacts of BDML on foresight practices already adapting to technological advancements. All projections were assessed in three aspects: expected probability of projection's occurrence (EP), desirability of projection's occurrence (DE) and impact on foresight industry if the projection occurs (IF). The aspects were measured on a five-point Likert scale, with the categories: 1) very low; 2) low; 3) neither low nor high; 4) high; 5) very high (Likert, 1932).

3.2.2 Online survey design. Supported by Survey Monkey platform, the online survey was structured in four parts: introduction; instructions; projections; and demographic questions. The introduction presented a succinct abstract, the research's objectives, the participant's contribution and confidentiality and privacy terms. The main concepts were explained in the instructions part (foresight, uncertainty, big data and machine learning), as well as the structure of the questions and projections. The next section presented the set of projections,

including dedicated open comment boxes for the projections in each stage of the foresight process. The last part of the survey was composed of demographic questions. To define the target audience, we follow the survey guidelines of Mota *et al.* (2020). The experts were invited according to the list of the corresponding authors' names and e-mails identified on the FSTI database. To increase the response rate, author's name, title of the paper and publication year were included in the invitation mail (Sauermann and Roach, 2013). The survey was also sent to the World Futures Studies Federation (WFSF) [5] and Millennium Project (MP) [6] members, through group mail. In total, 7,753 foresight experts were directly invited through e-mail via Survey Monkey platform. Descriptive statistics were used to analyze the variables related to the projections. Subsequently, the projections were grouped based on their average EP and IF scores, allowing us to identify patterns and prioritize implications for the foresight field. Finally, the software Microsoft Excel and Vantage Point were used to create tables and charts for data visualization.

4. Results

4.1 Bibliometric analysis of foresight studies supported by big data and machine learning
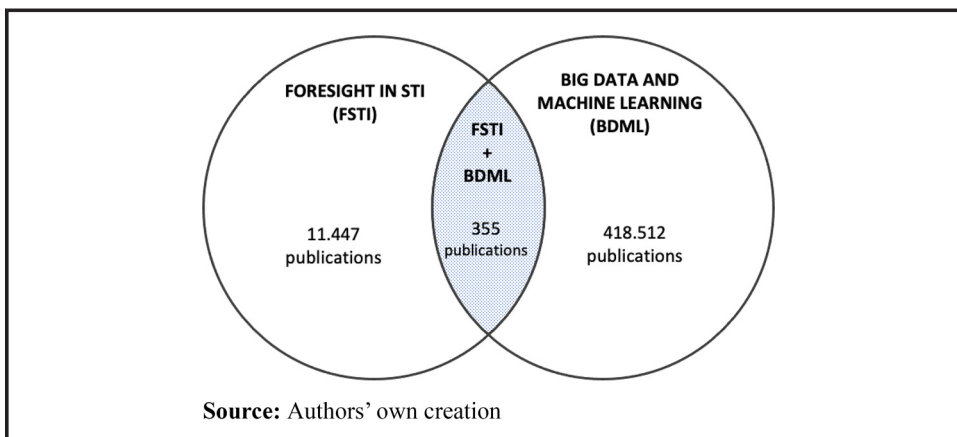
The publications were collected in November 2022. As is shown in Figure 2, the primary data corpus used to perform the analysis is based on foresight studies supported by BDML (FSTI+BDML). Figure 3 shows the annual distribution and the accumulated number of publications for the past 20 years, with an average of 14 publications per year. Although the number of publications has increased year upon year, research regarding the application of BDML in foresight is still incipient.

The most frequently used methodology (Figure 4) is patent analysis – a quantitative method that uses statistical methods to analyze patent data – which is mentioned in 21% of total papers. Text mining corresponds to 20%, followed by machine learning algorithms, which includes techniques such as linear and logistic regression, K-means clusterization, SVM, naïve Bayes, decision trees and random forest algorithms.

A foresight study can combine an average of six different methods (Popper, 2008); therefore, Figure 5 shows the co-occurrence network of foresight methodologies and the BDML techniques. Each node in this network represents a different approach (a foresight method or BDML technique), the size of the node represents the number of papers that use this method or technique and the colors of the nodes distinguish the type of the approach.
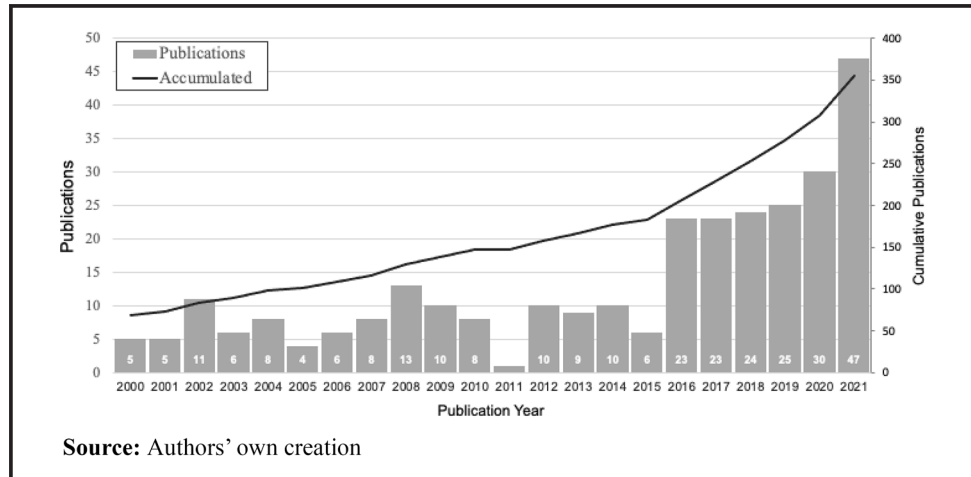
According to the network analysis, patent analysis is often used along with text and data mining techniques, machine learning algorithms and statistical methods. The network also

**Figure 2** Model of data collection results for FSTI, BDML and FSTI+BDML (1968–2021)

FORESIGHT IN STI (FSTI)

BIG DATA AND MACHINE LEARNING (BDML)

FSTI + BDML

11.447 publications

355 publications

418.512 publications

Source: Authors' own creation

Figure 3 Publications by year in FSTI+BDML

Source: Authors' own creation

Figure 4 Publications per methodological approach in FSTI+BDML papers


Source: Authors' own creation
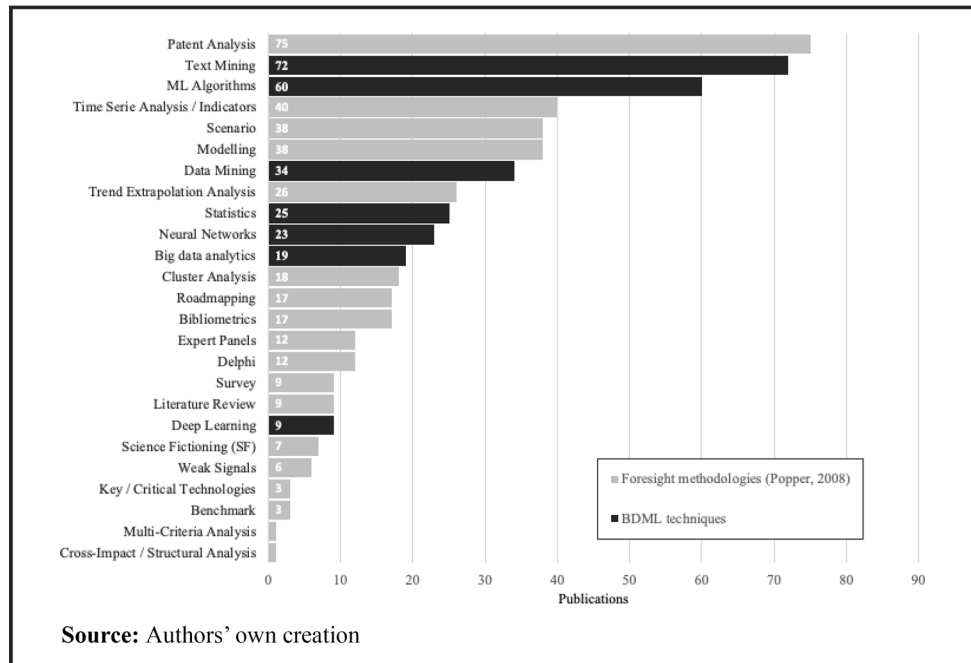
shows that foresight methodologies such as time series analysis, trend extrapolation analysis, modeling, roadmap and scenarios are often complemented with BDML techniques.

4.2 Survey assessment of future projections

The survey was available to receive answers from October 19th to October 26th, 2020. A total of 479 researchers and foresight experts (mostly authors from the collected bibliometric data) participated in this study. Most of participants were affiliated to universities (66,2%), followed by research institutions, consultancies and governmental institutions. Almost 60% of experts have declared more than ten years of experience in the futures field. Regarding the geographic region, half of consulted experts were from Europe (49.9%), followed by Asia

**Figure 5** Methodological approach's co-occurrence network in FSTI+BDML papers

Source: Authors' own creation
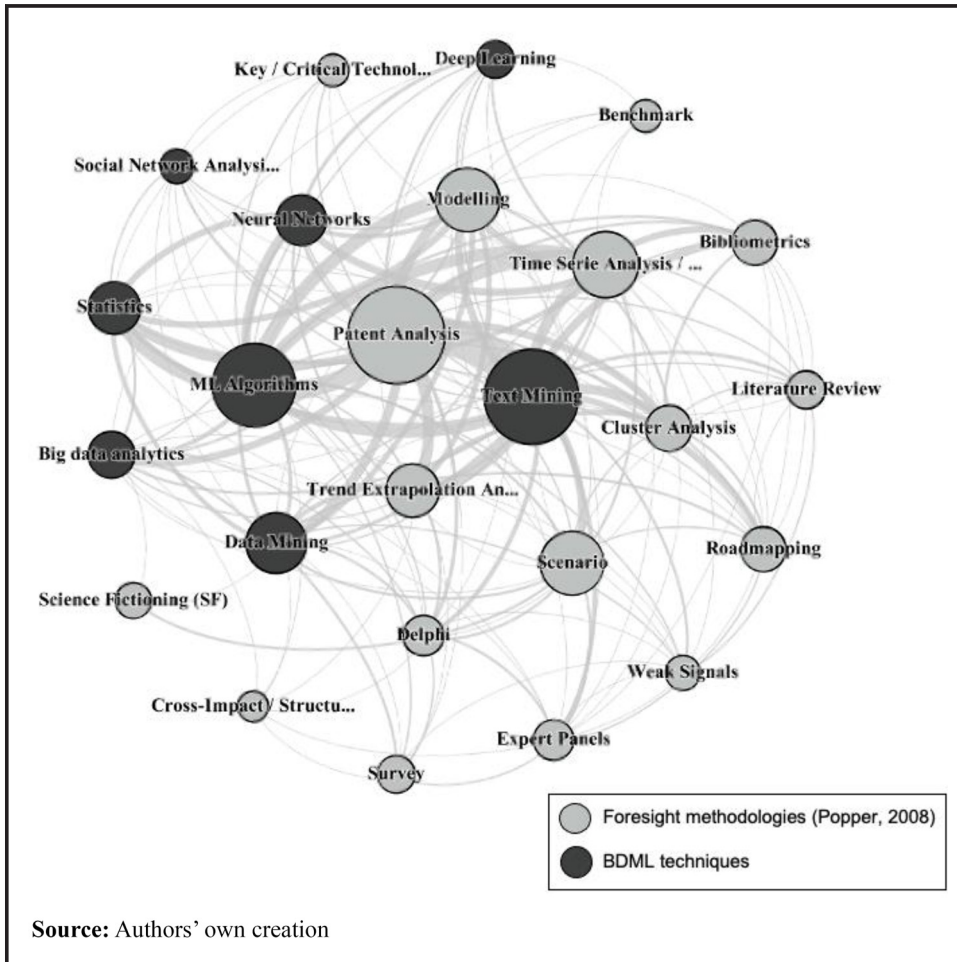
(15.9%), Northern America (14.2%) and Latin America and the Caribbean (13.9%). Figure 6 shows the distribution of answers and the mean value for each question and projection.

The projections P3 (require analytical skills) and P9 (support results transfer) were considered to have a higher EP. On the other hand, P1 (support objective definition) and P12 (support objectives' reach) were those with the lowest aggregate EP values. Some respondents expressed skepticism about the outcomes of BDML in foresight, citing variations in data quality and potential for introducing more bias. This is in line with low value of DE on P8 (increase data manipulation). Experts recognize data manipulation as a risk to data-driven foresight, potentially producing biased outcomes. The projections with the highest IF values were P3 (require analytical skill) and P4 (support data collection). The respondents expressed concerns on developing new competences to leverage BDML outputs. The general perception was that having dealt with the potential risks, BDML would improve data analysis in foresight, not necessarily reducing uncertainty.

In Figure 7, the projections were plotted with the mean value of IF on the *y*-axis, the mean value of EP on the *x*-axis and the mean value of DE as a proportion of the bubbles' size.

Projections were split into four different groups based on EP and FI. Group 1 is composed of projections with high EP and high IF, and includes the P3, P9 and P8. Group 2 has medium EP but still high IF, including data-related projections P4, P5, P6 and P7. With medium EP and

Figure 6 Distribution of answers in EP, DE and IF dimensions

Source: Authors' own creation
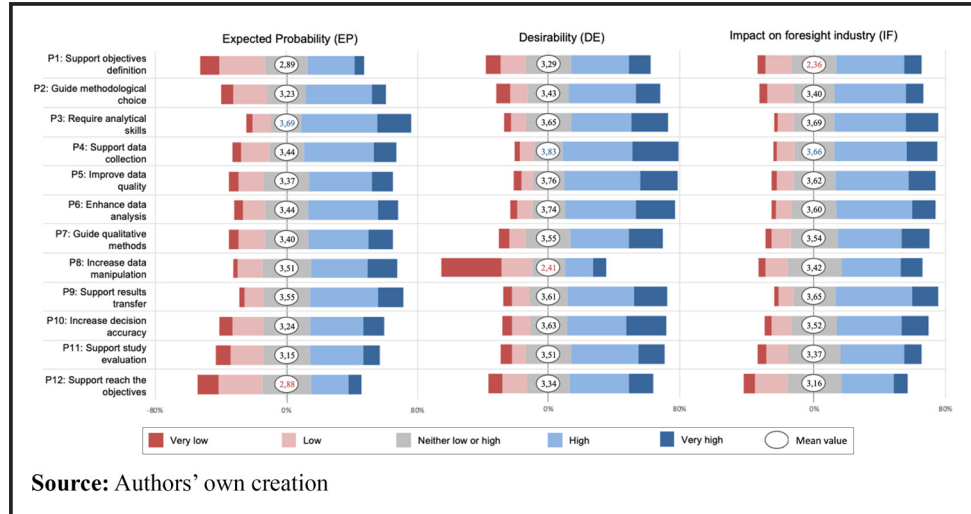
Figure 7 Group of projections based on EP and IF

Source: Authors' own creation
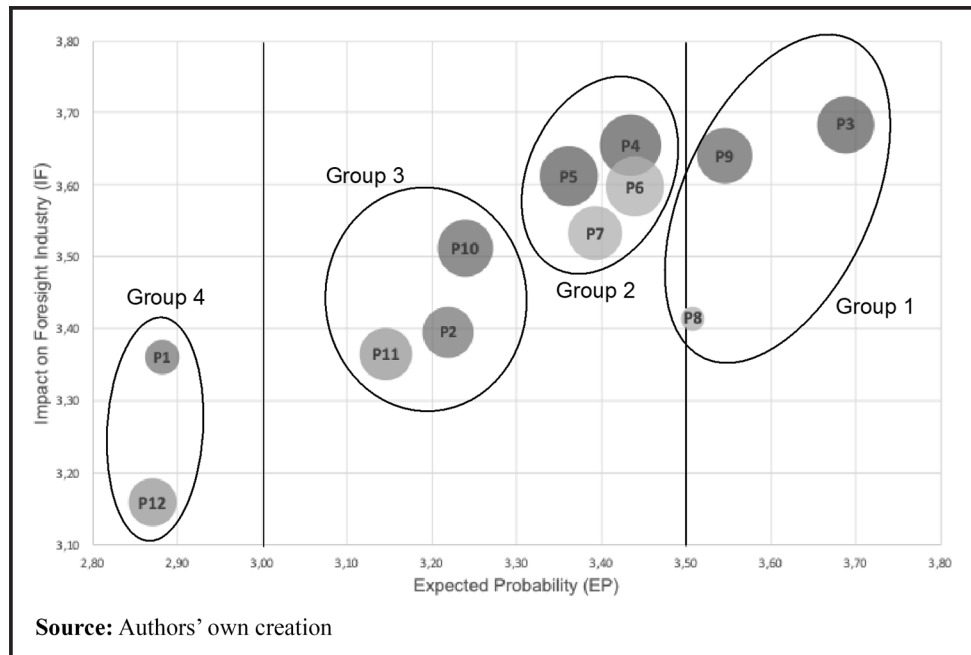
relatively low IF, Group 3 includes P2, P10 and P11. Group 4 presents low EP and IF and includes P1 and P12.

Some methodological limitations of the survey analysis must be mentioned. First, we could not run a complete comparison between the characteristics of the original sample (7,753 invited foresight experts) to those who gave complete and valid answers to the survey (479 respondents). However, the minimum sample size was achieved for 95% of confidence level and 5% of margin of error.

5. Discussion

Bibliometric data and experts' opinions demonstrate that foresight will be increasingly implemented and supported by BDML. This development will have two key impacts, over how foresight methodologies are developed and implemented and over the way practitioners will perceive and face uncertainty in decision-making and unfold foresight results to policymaking.

5.1 Effects on foresight methodologies and practice

The integration of BDML with prospective methodologies has emerged as a promising avenue, with the potential to attract new researchers, increase futures literacy and change the landscape of future studies (Spaniol and Rowland, 2023). Bibliometric results indicate a significant increase of BDML techniques within foresight methodologies, especially text and data mining (for pattern recognition and relationship analysis among concepts) and varied machine learning algorithms (for classification, predictive analytics and the recent generation of text and images through LLM). However, to be able to leverage the potential and mitigate risks of BDML in foresight, it is indispensable analytical competencies within foresight teams, as derived from high EP in P3. Furthermore, it is crucial to equip practitioners with a diverse skill set encompassing data literacy, data analytics, qualitative approaches and domain expertise, to deal with the increasing complexity of the future (Gary and von der Gracht, 2015; Tetlock and Gardner, 2015). Such blended competencies also facilitate the communication between stakeholders, enhancing participation, improving futures literacy and ensuring the quality of processes and results in foresight (Keller and von der Gracht, 2014). To this end, data scientists can contribute significantly by handling data collection, storage and cleaning, while providing accurate interpretations of analyses, in collaboration with practitioners and domain experts (Pankratova and Savastiyanov, 2014; Thu *et al.*, 2022).

The high EP in P9 suggests that BDML tools could be relevant in the implementation of foresight results by continuously monitoring trends, setting priorities, making decisions and supporting policy making (van Belkom, 2020; Keller and von der Gracht, 2014). Short-term automated decision-making is already a reality in some forecasting exercises (Chaboud *et al.*, 2014; Roozbehani *et al.*, 2010), but for foresight and long-term decisions (open and complex systems), human participation will still be required (Geurts *et al.*, 2022; Ivanov, 2022). In this context, Keenan *et al.* (2020) describe how digital tools can strengthen foresight and support the development and improvement of STI policies. As an example, Japan's National Graduate Institute for Policy Studies developed the SciREX Policymaking Intelligent Assistance System (SPIAS). SPIAS uses big data and semantic technologies to analyze socio-economic impacts of research, evaluate scientists' performance pre- and post-grants and map emerging technologies. By providing up-to-date data analysis, based on national and international databases, the system supports evidence-based policymaking and policy analysis in Japan.

The results on projections in Group 2 (P4, P5, P6, P7) highlighted the potential of data-driven analyses to complement and enhance existing foresight methodologies, as demonstrated in bibliometrics, including patent analysis, time-series analysis, scenarios and roadmaps. The mix of methods used in foresight studies, as illustrated in Figure 5, clearly demonstrates how BDML techniques are integrated with both quantitative and qualitative methodologies. Many authors (Denter *et al.*, 2022; Kayser *et al.*, 2014; Kayser and Shala, 2020; Nazarenko *et al.*, 2021; Santo *et al.*, 2006; Yufei *et al.*, 2016) describe examples of how BDML could support foresight practice and decision-making, by providing up-to-date information, guiding creative processes and potentially avoiding expert biases.

Despite the potential benefits and the increasing use of BDML in foresight, manipulation of data and biased algorithms emerge as challenges to practitioners and data scientists (P8) (L'Heureux *et al.*, 2017). Efforts to mitigate bias involve data management processes and the collaborative integration of human expertise with data analysis outcomes, to critically examine and contextualize data, as the hybrid AI-expert approach proposes (Geurts *et al.*, 2022).

BDML's influence on the initial (pre-foresight) and final (renewal) stages of the foresight process is expected to be limited, according to the results on P1, P2, P11 and P12. In these stages, tasks are predominantly shaped by other factors such as the sponsor's strategy, scope, costs, available competences and time for conducting the foresight study.

5.2 Effects on uncertainty and decision making

The results of P10, P12 and projections on Group 2 suggest that the use of BDML is neither expected to simplify the complexity of foresight nor the accuracy of decision-making under uncertainty but may modify the "partiality" of knowledge. For the time being, knowledge will remain partial when referring to the future and will certainly change the bases on which uncertainty will be delimited by agents, in the sense proposed by Shackle (1969) and Metcalfe et al. (2021).

The premise that "the more data and information on a given subject, the easier it will be for a better decision" depends on how the subject is framed. For well-defined subjects that are less affected by the emergence of new variables – a condition of risk and ambiguity or uncertainty levels 1 and 2 – more data can narrow the possibilities that lie ahead. There is evidence that short-term, data-driven forecast studies will be the most impacted by the adoption of BDML tools, in line with the empirical findings of Tetlock and Gardner (2015). On the other hand, BDML can also open more prospecting fronts in open systems (possibly revealing new variables) and even increasing the complexity of decision-making, by modifying the partiality of knowledge.

BDML applied to futures studies may only interfere over contexts of uncertainty if they are able to convert uncertainty into ambiguity or risk. In other words, converting open systems into closed ones. This would require encompassing all potential future possibilities for a specific context within a certain timeframe. For it to be observable, this situation would have to endure for at least the time it was predicted for. In this condition, either time would not act to bring about change or BDML tools would be able to anticipate any changes, making that system a closed one for a certain timeframe.

The approach that Marchau et al. (2019) put forth is particularly valuable at this juncture. The authors' suggestion to "split" uncertainty into 4 + 1 levels allows us to inquire whether BDML will transform contexts characterized by Levels 3 (a few plausible futures) and 4a (many plausible futures) into Level 2 (stochastic system model), or into Level 1 (single deterministic system model). The prospect of BDML transforming a state of unknown futures (Level 4b) into Level 2 or 1 appears even more challenging. This is because it would necessitate an even greater reliance on its generative capability to imagine futures autonomously. Theoretically speaking, all these transformations are possible. The extent to which this will occur cannot yet be calculated nor proved. Again, it will depend on the capacity of BDML to close systems, which can only be checked in the future.

6. Final remarks

This study aims to identify conceptual and methodological implications of integrating BDML into the field of foresight. While the adoption of BDML in foresight remains in its nascent stages, its presence is expanding, particularly in academic literature. Foresight experts and practitioners anticipate some effects from BDML adoption. Primarily, it is growing the demand of novel competencies among foresight team, such as data literacy and data analytics skills, to promote a dynamic interaction between data outputs and human knowledge. These competencies, associated with proper institutional control mechanisms, possess the potential to mitigate latent risks associated with biased data and algorithms, thereby ensuring more reliability and robustness of insights derived from BDML-supported foresight. Second, foresight methods will be boosted by BDML techniques, which could enhance methodological capacity, automatically monitor societal and technological trends and provide qualified knowledge for prospective activities, opening path for imagining new and desirable futures. Consequently, foresight could

be extremely valuable for decision-making and evidence-based policy formulation, mainly for addressing societal challenges, such as climate change.

Uncertainty will always prevail in open systems, which means that, whatever the evolution of BDML (and other similar approaches as for LLM), any (big) data nor complex analysis will ever foresee the future with fully accuracy. Data can improve qualified information, but it will hardly replace imagination, creativity and expectation, which is essential for practitioners to imagine futures (van Belkom, 2020; Boysen, 2020; Geurts *et al.*, 2022) Rather, BDML's most important effect on prospective is precisely its influence on imagination, creativity and expectation. If this is true, the perception of uncertainty may even increase because large amounts of data and processing power will open new frontiers of knowledge and new alternatives futures. The parallel with the concept of event horizon helps to understand this point. The more details (data) are visible, the more one can know about what delimitates the event horizon at a given point in time. However, as the events on the horizon change continuously, the more you get, the more you will need to delimit it. Time and context always matter in future studies.

Notes

1. The ideas of partial knowledge and bounded uncertainty align with the well-established concept of bounded rationality put forth by Herbert Simon (Simon, 1959). The concept that human rationality is limited by the extent of available information and the various calculation methods used by economic agents is rooted in similar notions as those of Knight and Shackle.

2. For the time being, the idea of reversing the time's arrow (the symmetry of time) belongs to the domain of quantum thermodynamics, and it is a matter of the microscopic-cum-atomic level. At the macro level, the arrow of time is, to the best of our knowledge, one-way, and toward the future as proposed by the famous British astrophysicist Arthur Eddington.

3. This query is a combination of FSTI and BDML databases' queries: TITLE-ABS-KEY ("la prospective" OR "foresight" OR "technolog* forecast*" OR "technolog* anticipation" OR "technolog* prediction" OR "future oriented stud*" OR "future oriented analys*") AND TITLE-ABS-KEY ("science" OR "technolog*" OR "innovation") AND TITLE-ABS-KEY ("big data" OR "text mining" OR "data mining" OR "machine learning" OR "data analytics" OR "deep learning" OR "artificial intelligence") AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")).

4. Vantage Point is a text analytics software produced by Search Technology Inc., focused on analyzing scientific, technical, market and patent information.

5. WFSF is a UNESCO and UN consultative partner with members in over 60 countries. It is a forum for discussing ideas, visions and plans for alternative futures (for more information, access https://wfsf.org/).

6. Millennium Project is a global participatory think tank with 67 nodes around the world (for more information, access www.millennium-project.org/).

References

Albert, T., Moehrle, M.G. and Meyer, S. (2015), "Technology maturity assessment based on blog analysis", *Technological Forecasting and Social Change*, Vol. 92, pp. 196-209.

Amanatidou, E., Butter, M., Carabias, V., Konnola, T., Leis, M., Saritas, O., Schaper-Rinkel, P., *et al*. (2012), "On concepts and methods in horizon scanning: lessons from initiating policy dialogues on emerging issues", *Science and Public Policy*, Vol. 39 No. 2, pp. 208-221.

Boysen, A. (2020), "Mine the gap: augmenting foresight methodologies with data analytics", *World Futures Review, SAGE Publications*, Vol. 12 No. 2, pp. 239-248.

Brandtner, P. and Mates, M.A. (2021), "Artificial intelligence in strategic foresight – Current practices and future application potentials: current practices and future application potentials", ACM International Conference Proceeding Series, pp. 75-81.

Cabral, B., Fonseca, M. and Mota, F. (2019), "What is the future of cancer care? A technology foresight assessment of experts' expectations", *Economics of Innovation and New Technology*, Vol. 28 No. 6, pp. 635-652, doi: 10.1080/10438599.2018.1549788.

Chaboud, A.P., Chiquoine, B., Hjalmarsson, E. and Vega, C. (2014), "Rise of the machines: algorithmic trading in the foreign exchange market", *The Journal of Finance*, Vol. 69 No. 5, pp. 2045-2084.

Denter, N.M., Aaldering, L.J. and Caferoglu, H. (2022), "Forecasting future bigrams and promising patents: introducing text-based link prediction", *Foresight*.

Dequech, D. (2000), "Fundamental uncertainty and ambiguity", *Eastern Economic Journal*, Vol. 26 No. 1, pp. 41-60.

Dosi, G. and Egidi, M. (1991), "Substantive and procedural uncertainty – An exploration of economic behaviours in changing environments", *Journal of Evolutionary Economics*, Vol. 1 No. 2, pp. 145-168.

Ellsberg, D. (1961), "Risk, ambiguity, and the savage axioms", *The Quarterly Journal of Economics*, Vol. 75 No. 4, p. 643.

Gary, J.E. and von der Gracht, H.A. (2015), "The future of foresight professionals: results from a global Delphi study", *Futures*, Vol. 71, pp. 132-145.

Georghiou, L., Harper, J.C., Keenan, M., Miles, I., Popper, R. and Elgar, E. (2008), *The Handbook of Technology Foresight*, Edward Elgar Publishing, Cheltenham.

Geurts, A., Gutknecht, R., Warnke, P., Goetheer, A., Schirrmeister, E., Bakker, B. and Meissner, S. (2022), "New perspectives for data-supported foresight: the hybrid AI-expert approach", *Futures & Foresight Science*, Vol. 4 No. 1, p. e99.

Glassey, O. (2012), "Folksonomies: spontaneous crowd sourcing with online early detection potential?", *Futures*, Vol. 44 No. 3, pp. 257-264.

Hagen, L., Yi, H.S., Pietri, S. and Keller, T.E. (2019), "Processes, potential benefits, and limitations of big data analytics: a case analysis of 311 data from city of Miami", ACM International Conference Proceeding Series, pp. 1-10.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), "The elements of statistical learning", *The Mathematical Intelligencer*, Vol. 27.

Hess, D.J. (1997), *Science Studies: An Advanced Introduction*, NYU Press, New York, NY.

Irvine, J. and Martin, B.R. (1984), *Foresight in Science: Picking the Winners*, F. Pinter, London; Dover N.H.

Ivanov, S.H. (2022), "Automated decision-making", *Foresight*, doi: 10.1108/FS-09-2021-0183.

Karaca, F. and Öner, M.A. (2015), "Scenarios of nanotechnology development and usage in Turkey", *Technological Forecasting and Social Change*, Vol. 91, pp. 327-340.

Kayser, V. and Blind, K. (2017), "Extending the knowledge base of foresight: the contribution of text mining", *Technological Forecasting and Social Change*, Vol. 116, pp. 208-215.

Kayser, V. and Shala, E. (2020), "Scenario development using web mining for outlining technology futures", *Technological Forecasting and Social Change*, Vol. 156, p. 120086.

Kayser, V., Goluchowicz, K. and Bierwisch, A. (2014), "Text mining for technology Roadmapping - the strategic value of information", *International Journal of Innovation Management*, Vol. 18 No. 3, p. 1440004.

Keenan, M., Plekhanov, D., Galindo-Rueda, F. and Ker, D. (2020), "The digitalisation of science and innovation policy", *The Digitalisation of Science, Technology and Innovation: Key Developments and Policies*, OECD Publishing, Paris, pp. 165-182, doi: 10.1787/0fbe3397-en.

Kehl, W., Jackson, M. and Fergnani, A. (2020), "Natural language processing and futures studies", *World Futures Review* SAGE Publications CA, Los Angeles, CA, Vol. 12 No. 2, pp. 181-197.

Keller, J. and von der Gracht, H.A. (2014), "The influence of information and communication technology (ICT) on future foresight processes - results from a Delphi survey", *Technological Forecasting and Social Change*, Vol. 85 No. 2014, pp. 81-92.

Knight, F. (1921), *Risk, Uncertainty and Profit*, Houghton Mifflin Company, Boston; New York, NY.

L'Heureux, A., Grolinger, K., Elyamany, H.F. and Capretz, M.A.M. (2017), "Machine learning with big data: challenges and approaches", *IEEE Access*, Vol. 5, pp. 7776-7797.

Lecun, Y., Bengio, Y. and Hinton, G. (2015), "Deep learning", *Nature, Nature Publishing Group*, Vol. 521 No. 7553, 27 May.

Likert, R. (1932), "A technique for the measurement of attitudes", *Archives of Psychology*, Vol. 22 No. 140, p. 55.

Loukides, M. (2010), "What is data science? The future belongs to the companies and people that turn data into products", O'Reilly Radar, 2 June, available at: www.oreilly.com/radar/what-is-data-science/ (accessed 11 January 2021).

Loveridge, D. (2009), *Foresight: The Art and Science of Anticipating the Future*, Routledge.

Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M. and Popper, S.W. (2019), *Decision Making under Deep Uncertainty*, Springer International Publishing, Cham, doi: 10.1007/978-3-030-05252-2.

Metcalfe, S., Salles-Filho, S., Duarte, L.T., Bin, A., Azevedo, A.T. and Feitosa, P.H.A. (2021), "Shackle's approach towards priority setting and decision-making in science, technology, and innovation", *Futures*, Vol. 134, p. 102838.

Miles, I. (2002), "Appraisal of alternative methods and procedures for producing regional foresight", Conference: European Commission's DG Research Funded STRATA – ETAN Expert Group Action, *CRIC, Manchester, UK*, p. 21.

Miles, I. (2010), "The development of technology foresight: a review", *Technological Forecasting and Social Change*, Vol. 77 No. 9, pp. 1448-1456.

Miles, I., Harper, J.C., Georghiou, L., Keenan, M. and Popper, R. (2008), "The many faces of foresight", *The Handbook of Technology Foresight-Concepts and Practices*.

Miles, I., Meissner, D., Vonortas, N.S. and Carayannis, E. (2017), "Technology foresight in transition", *Technological Forecasting and Social Change*, Vol. 119 No. April, pp. 211-218.

Miles, I., Saritas, O. and Sokolov, A. (2016), "Foresight for science, technology and innovation", doi: 10.1007/978-3-319-32574-3.

Mittelstadt, B.D. and Floridi, L. (2016), "The ethics of big data: current and foreseeable issues in biomedical contexts", *Science and Engineering Ethics*, Vol. 22 No. 2, pp. 303-341.

Mota, F., Braga, L., Rocha, L. and Cabral, B. (2020), "3D and 4D bioprinted human model patenting and the future of drug development", *Nature Biotechnology*, Vol. 38 No. 6, pp. 689-694.

Nazarenko, A., Vishnevskiy, K., Meissner, D. and Daim, T. (2021), "Applying digital technologies in technology roadmapping to overcome individual biased assessments", *Technovation*, Vol. 110, p. 102364.

Pang, A.S.K. (2010), "Social scanning: improving futures through web 2.0; or, finally a use for twitter", *Futures*, Vol. 42 No. 10, pp. 1222-1230.

Pankratova, N. and Savastiyanov, V. (2014), "foresight process based on text analytics", *Information Content and Processing*, Vol. 1 No. 1, pp. 54-65.

Popper, R. (2008), "How are foresight methods selected?", *Foresight*, Vol. 10 No. 6, pp. 62-89.

Prigogine, I. and Stengers, I. (1984), "Order out of chaos", *Man's New Dialog with Nature*, Famingo Edition, London.

Reimsbach-Kounatze, C. (2015), "The proliferation of 'big data' and implications for official statistics and statistical agencies", *OECD Digital Economy Papers*, No. 245, pp. 3-39, doi: 10.1787/5js7t9wqzvg8-en.

Roozbehani, M., Dahleh, M. and Mitter, S. (2010), "Dynamic pricing and stabilization of supply and demand in modern electric power grids", 2010 1st IEEE International Conference on Smart Grid Communications, SmartGridComm 2010, pp. 543-548.

Rožanec, J., Nemec, P., Leban, G. and Grobelnik, M. (2023), "AI, what does the future hold for US? Automating strategic foresight", *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering*, pp. 247-248, doi: 10.1145/3578245

Santo, M., Coelho, G.M., dos Santos, D.M. and Filho, L.F. (2006), "Text mining as a valuable tool in foresight exercises: a study on nanotechnology", *Technological Forecasting and Social Change*, Vol. 73 No. 8, pp. 1013-1027.

Saritas, O., Burmaoglu, S. and Ozdemir, D. (2022), "The evolution of foresight: what evidence is there in scientific publications?", *Futures*, Vol. 137, p. 102916.

Sauermann, H. and Roach, M. (2013), "Increasing web survey response rates in innovation research: an experimental study of static and dynamic contact design features", *Research Policy*, Vol. 42 No. 1, pp. 273-286.

Shackle, G.L.S. (1969), *Decision Order and Time in Human Affairs*, Cambridge University Press, Cambridge, pp. 3-113.

Shackle, G.L.S. (2010), *Uncertainty in Economics and Other Reflections*, reprint, Cambridge University Press.

Shalev-Shwartz, S. and Ben-David, S. (2014), *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.

Simon, H.A. (1959), "Theories of decision-making in economics and behavioral science", *The American Economic Review*, Vol. 49 No. 3, pp. 253-283.

Spaniol, M.J. and Rowland, N.J. (2023), "AI-assisted scenario generation for strategic planning", *Futures & Foresight Science*, Vol. 5 No. 2, p. e148.

Stahl, B.C. and Wright, D. (2018), "Ethics and privacy in AI AND big data: implementing responsible research and innovation", *IEEE Security & Privacy*, Vol. 16 No. 3, pp. 26-33, doi: 10.1109/MSP.2018.2701164.

Tetlock, P.E. and Gardner, D. (2015), *Superforecasting: The Art and Science of Prediction., Superforecasting: The Art and Science of Prediction*, Crown Publishers/Random House, New York, NY, US.

Thu, M.K., Beppu, S., Yarime, M. and Shibayama, S. (2022), "Role of machine and organizational structure in science", *Plos One*, Vol. 17 No. 8, p. e0272280.

Trappey, A.J.C., Trappey, C.V., Govindarajan, U.H. and Sun, J.J.H. (2019), "Patent value analysis using deep learning models - the case of IoT technology mining for the manufacturing industry", *IEEE Transactions on Engineering Management*, Vol. 68 No. 5, pp. 1-13.

van Belkom, R. (2020), "The impact of artificial intelligence on the activities of a futurist", *World Futures Review, SAGE Publications*, Vol. 12 No. 2, pp. 156-168.

von der Gracht, H.A., Bañuls, V.A., Turoff, M., Skulimowski, A.M.J. and Gordon, T.J. (2015), "Foresight support systems: the future role of ICT for foresight", *Technological Forecasting and Social Change*, Vol. 97, pp. 1-6.

Xu, L.D., Xu, E.L. and Li, L. (2018), "Industry 4.0: state of the art and future trends", *International Journal of Production Research*, Vol. 56 No. 8, pp. 2941-2962.

Yufei, L., Yuan, Z. and Ling, L. (2016), "Application of big data analysis method in technology foresight for strategic emerging industries", *Chinese Journal of Engineering Science*, Vol. 18 No. 4, p. 121.

Zhou, Y., Dong, F., Liu, Y., Li, Z., Du, J.F. and Zhang, L. (2020), "Forecasting emerging technologies using data augmentation and deep learning", *Scientometrics*, Vol. 123 No. 1, pp. 1-29.

About the authors

Vinicius Muraro is a Postdoctoral Fellow and full member at CIRCLE – Center for Innovation Research, at Lund University (Sweden). He holds a PhD in Science and Technology Policy from the University of Campinas (Unicamp – Brazil) and has eight years of professional experience in innovation consulting, focusing on impact evaluation and technology foresight for large Brazilian companies and research institutes. His research interests are innovation management science of science, foresight for STI, artificial intelligence and quantitative methods. Vinicius Muraro is the corresponding author and can be contacted at: murarosilva@gmail.com

Sergio Salles-Filho is a Full Professor in the Department of Science and Technology Policy at the University of Campinas (Unicamp – Brazil). He was director of the Faculty of Applied Sciences at Unicamp from 2010 to 2013 and director of the Institute of Geosciences at the same university between 2017 and 2021. He was visiting researcher at the Manchester Institute of Innovation Research between 2013 and 2014.