An Exploration of the Potential of Machine Learning Tools for Media Analysis to Support Sense-Making Processes in Foresight

Michela Vignoli[1][*][†], Jan Rörden[2][†], Dana Wasserbacher[3][†] and Simone Kimpeler[2][†]

[1] AIT Austrian Institute of Technology, Center for Digital Safety & Security, Vienna, Austria, [2] Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, Germany, [3] AIT Austrian Institute of Technology, Center for Innovation Systems & Policy, Vienna, Austria

In view of the many discussions about uncertainty regarding the further development of the coronavirus disease 2019 (COVID-19) pandemic and its effects on the economy and society, we observed that the crisis led to an increased presence of individual researchers and experts making forward-looking statements on the impacts of the COVID-19 pandemic or stating trends in mass media publications. From a strategic foresight research perspective, there is a need to further analyse an increase of future-oriented expert statements in public media in a context of high uncertainty like the impacts of the COVID-19 pandemic and related crises. Given the increasing amount of media texts available for web-based scanning and text analysis, Machine Learning (ML) is a promising approach for text analysis of big data, which also raises high expectations in the field of foresight, particularly in the context of scoping and scanning activities for weak signal detection and text analysis for sense-making processes. In this study, we apply a natural language processing (NLP)-based ML approach to analyse a large corpus of news articles from web sources to explore the potential of applied ML to support sense-making in the field of foresight, specifically for the analysis of future-related statements or predictive statements in media. The results underline the potential of ML approaches as a heuristic tool to support sense-making in foresight processes and research, particularly by pre-structuring large datasets (e.g., news articles around a particular topic of public debate). The ML can provide additional insights for actor analysis associated with a specific topic of public debate from a large data corpus. At the same time, our results show that ML models are limited in their ability to provide solid evidence and that they can also lead to fallacies. Therefore, an ML can only be considered as a heuristic tool supporting specific steps in a sense-making process and development of further research questions, as well as encouraging reflection on the application of ML-based approaches in foresight.

Keywords: science communication, foresight, Machine Learning, news media, big data, open science, COVID-19

INTRODUCTION

Since the start of the coronavirus disease 2019 (COVID-19) pandemic, the presence of scientific experts from various research fields has increased sharply in public media (Ioannidis et al., 2021), social media (Ahmed et al., 2020; Chipidza et al., 2021), and the news (Krawczyk et al., 2021). The COVID-19 pandemic is described as the most critical science communication challenge (O'Connor et al., 2021). These developments are also a new methodological challenge for the practice field of strategic foresight, as forward-looking statements are gaining importance in the context of scientifically based statements about the medium- and long-term effects of the COVID-19 pandemic on society and the economy. For foresight research and future analysis, two aspects, in particular, are exciting from our point of view: On the one hand, it can be expected that this will increase the number of future-oriented statements and even predictions of the future in general, as well as the breadth of scientific expertise with regard to future statements in the thematically focused area of the consequences of the pandemic; while on the other hand, this increase should also make it possible to obtain a larger number of future statements from different research areas, which can be evaluated in terms of content for various activities within the framework of strategic foresight, e.g., scenario development, or the derivation of policy options. The latter is the so-called sense-making phase, in which the candidate signals for change are evaluated in terms of their relevance to concrete issues and needs for action. The sense-making in foresight processes can be defined as a cognitive process that supports the dealings with uncertainties and the unknown in decision making, and structures the unknown by linking signals of change to better understand the complex cross-impacts and to inform the decision-makers.

To further investigate this hypothesis, specifically how the sense-making based on cognitive processes could be supported by the evaluation of larger amounts of data with a semi-automated Machine Learning (ML) approach, we chose a public news media data corpus to search for insights and evidence for the increase in forward-looking statements by actors in the news and public media from different research fields since the outbreak of the COVID-19 pandemic. The goal was to explore the potential of such an ML approach to identify evidence in a large data corpus of news articles for sense-making processes in foresight endeavours.

Starting with a scanning activity, we used text mining and a natural language processing (NLP) model within an ML approach to identify forward-looking statements related to COVID-19 in a large data corpus of online news articles. In the next step, we tried to make sense out of the identified statements (assessment of the relevance and novelty of the content) and again applied an ML approach to gain further insights from the context of the identified statements (such as information on the actors associated with the content). In the last step, we tried to integrate our different and interdisciplinary learning experiences from applying ML into a strategy for future research by highlighting challenges and opportunities of ML to support sense-making processes in foresight. While Geurts et al. (2021)

already provided evidence for the usefulness of ML in the horizon scanning phase, the question remains as to how far ML can contribute to sense-making and strategy development activities.

In this article, we follow a systemic foresight approach, which builds on horizon scanning and sense-making methods to explore possible futures and apply ML specifically for improving sense-making activities by analysing large amounts of data that contain future-oriented statements in the context of the COVID-19 pandemic. We present the first results of an empirical study based on a semi-automated method to identify and extract forward-looking statements and names of actors mentioned in the context of the statements from the full texts of online news articles. The chosen methodological approach, as well as the underlying concepts and definitions, are explained in more detail in Sections Materials and Methods and Methodological Approach.

Results of this first attempt to apply ML techniques for identifying evidence for foresight knowledge are included in Section Results: ML Classification of Forward-Looking Statements.

In the final chapter, we conclude to which extent our proposed approach has the potential to provide an evidence base for identifying forward-looking statements and related actors in news articles. Limitations of the chosen approach, such as 1) limited adequacy and quality of the used datasets; and 2) limitations to a generic working definition of 'forward-looking statements', are discussed as well. In Section Discussion, we discuss challenges and opportunities that the applied ML techniques offer in the context of foresight practice, as well as potential future research directions.

MATERIALS AND METHODS

What Is Foresight?

Foresight is a systemic approach for strategic and forward-looking activities of exploring multiple futures (European Commission, 2020) in a structured way and is a broad field of research and practice at the interface of policy and science, which aims at generating forward-looking knowledge and enhancing anticipatory capacities to support decision-making and dealing with uncertainties (Robinson et al., 2021). In many foresight processes, it is key to involve experts, stakeholders, and citizens in creative sense-making workshops and strategic dialogues about the future of complex issues (or complex futures). It is important to understand that foresight is 'not about predicting the future but about exploring different plausible futures that could arise and the opportunities and challenges they could present' (European Commission, 2020). In this sense, strategic foresight is different from forecasting. While forecasting attempts to predict a single 'correct' version of the future based on data from the past, evidence, and probability (e.g., mathematical modelling), a foresight uses multiple alternative plausible futures based on plausible combinations of various possible developments of influencing factors to identify future risks and challenges under uncertainty. Besides its systematic, participatory, future-intelligence-gathering character, the more qualitatively oriented foresight often aims at medium- to long-term vision-building

process for present-day decisions or mobilisation of joint actions (Gavigan et al., 2001, p. 3). In the latter definition, the emphasis lies on the interaction of multiple actors involved in the process of generating and deliberating foresight knowledge and overcoming individual and institutional filters of perception and biases of future assumptions (Warnke and Schirrmeister, 2016; Rosa et al., 2021). Both quantitative and qualitative methods are often combined in foresight processes, and with the increasing digitisation and availability of big data, the interest in deploying semi-automated or automated methods to explore large amounts of data is also growing in this field (cf. Geurts et al., 2021).

The architectures and forms of implementation of foresight processes in organisations vary widely, depending on the specific objective and context of the foresight. For this study, we refer to a framework that identifies three distinctive activities in a full-cycle foresight process: 1) Horizon scanning to identify weak signals of change in society, technology, economy, ecology, and politics (STEP); 2) Insights and sense-making activities to reflect on cross-impacts between the trends and drivers, and develop a variety of plausible scenarios and visions of preferred futures; 3) Strategy development by analysing scenarios, strength, and weaknesses or gaps between actions needed and current strategies; In some contexts, in particular, in strategic foresight for organisations, a fourth phase could be added; 4) Implementation and action, to start the actions, which could be outlined in a roadmap or strategic plan, and monitor its achievements (Cuhls et al., 2015; see **Figure 1**: Foresight cycle). It is important to note that these activities are not separated steps in the process, but rather interrelated in practice. For example, the transition between horizon scanning and sense-making is fluid, and the strategy development phase may again involve scanning and participatory sense-making methods.

FIGURE 1 | Foresight cycle.
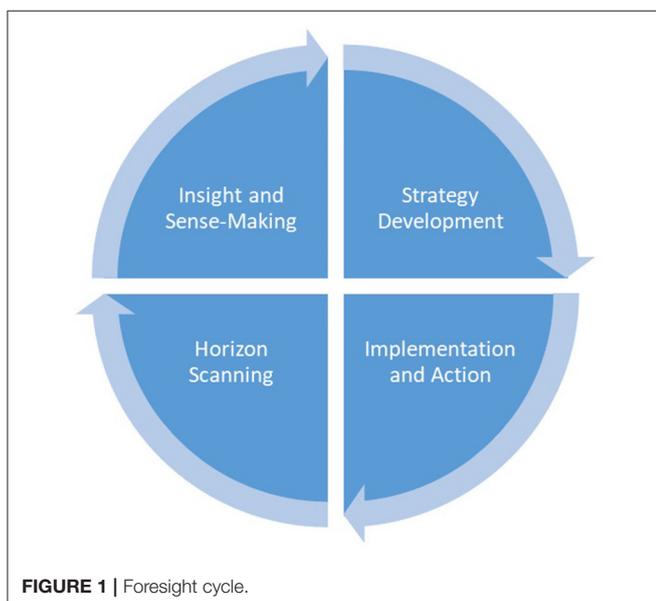
Foresight approaches, which are engaging and using the collective knowledge of larger communities or the public, vary widely[1]. An open concept that is extensively applied in foresight practice and is subject to constant deliberation is relevant to contextualising our empirical study: the 'futures literacy'. The UNESCO has successfully established the futures literacy approach, with chairs in foresight around the globe, promoting futures literacy as a capacity to analyse futures-related information and insights and to use futures (images, scenarios, and narratives of possible futures) in today's decision-making (Miller, 2018). Since the future is uncertain, climate change, pandemics, or economic crisis challenge the individual and the collective images and narratives of the future that are used to make decisions today; the knowledge about the future, or 'foresight knowledge' mainly depends often on the non-reflected practices used to generate visions. An assessment of imaginaries and narratives of the future in a society (Jasanoff and Kim, 2015), as shared, for instance, in public media discourses, must take, in particular, the account of the content and the individual and group sense-making processes involved.

Therefore, strategic foresight is decisively dependent on the scanning of content and text data, and the analysis of spotted signals of change. Furthermore, the combination of different stakeholder and scientific expert perspectives to make sense out of the data is decisive in the context of the strategic decision that is considered the best or most relevant knowledge on a specific issue (Loveridge, 2004, p. 46). Frequently, the foresight exercises, which are based on scientific expertise, do not make the selection process transparent and seldom examine the elicitation process of expert opinion (Loveridge, 2004, p. 39). This tension opens a normative dimension addressing the qualification of the available information, which is especially relevant when communicating the results of a foresight exercise to decision-makers (Von Schomberg et al., 2006, p. 149).

The Potential of Applied Machine Learning for Foresight

Machine Learning (ML) is not a new research discipline (McCarthy et al., 1955) but is increasingly often introduced into new domains. This is done for various reasons, and amongst them is an improved pattern recognition, which fueled the recent rise in ML applications and can make use of labelled or unlabeled data (LeCun et al., 2015). If combined with suitable computing resources, it creates the opportunity to process large amounts of data, and extract information from it, thus, both expanding the amount of data available for additional analysis, as well as countering biases, and, generally, supporting decision making. The combination of human expert knowledge and algorithms that can solve lengthy and often tedious tasks and integrate vast amounts of data is of interest for various domains, i.e., in foresight (c.f. Geurts et al., 2021), where it can be integrated into each step of the foresight process. Within ML, several generalised methods can be used for interdisciplinary work. With a focus on the work presented here and the usage for text-based research, the

---

[1]For an overview on foresight and foresight methodology see Georghiou et al. (2008) or Loveridge (2009).

NLP algorithms are a specialised family of algorithms that can be used for many tasks. Additionally, more generalised techniques, such as clustering algorithms, are used here.

Assessing future-related evidence in a large data set needs a classification model for forward-looking statements. We propose two sets of elements to focus the search field of the horizon scanning (see Section Identification and Classification of Forward-Looking Statements). We developed the elements based on experiences from previous horizon scanning research projects, i.e., Foresight Fraunhofer I and II[2].

How to Extract Evidence From a Large Data Corpus of News Articles to Support Foresight Practice?

Both Foresight and ML heavily depend on sense-making activities that aim at transforming a complex body of data to information and knowledge, to make it manageable, workable, communicable, and discussable. They demark processes, in which knowledge is alternately and continuously collected, structured, rearranged, enriched, and summarised. Especially the main characteristic of foresight knowledge, which is the uncertainty, makes it necessary to conceive of assessment as an open and ongoing process aimed at making explicit, and continuously reflecting on the implicit assumptions that frame relevant forward-looking statements.

Muller et al. (2016) noted that ML should be considered an iterative process occurring at multiple and nested levels rather than a linear process. They observed that the 'dialogic iteration between human researcher and computational algorithm' is often not considered in ML applications. However, in our study, we found that human interaction and judgement played a central role, particularly in the context of the iterative classification process that led to the creation of our ground-truth data.

It should be noted that the goal of ML for problem-solving and reduction of complexity lies primarily in the quantity, which is the recognition of patterns in a sheer and unmanageable amount of data. Foresight processes are increasingly defined as participatory processes to take account of diverse stakeholder perspectives and collective intelligence and, therefore, involve stakeholders, citizens, and scientific experts from different disciplines and domains in their sense-making processes. In contrast, ML models are often developed single-handedly by computer scientists with specialised training in data analysis and modelling. The ML techniques are based on learning from their own inherent models that are happening in the notorious 'black box' of ML, and the reflexive analysis with different perspectives and creation of join knowledge from a variety of actors does not happen often. To a certain extent, then, the trap of expert-centrism and its lack of transparency with regard to the negotiation of expert opinions is as inherent to ML techniques as it is to expert-centred foresight processes.

Ultimately, the classification process of the ML model cannot be explained for every single decision, unless additional work is carried out for this specific data set, potentially at the cost

of the model performance. The reasons for similarity scores are not made transparent by design, so, it cannot be said why the ML algorithm classified some text snippets with a high similarity score, while others are not. Although we did not further explore this within our work, there are approaches and techniques aimed at explaining classification processes of ML to render the black box more transparent [Explainable AI/XAI, for example, lime, as presented by Ribeiro et al. (2016)]. On the one hand, this may seem like a radical limitation of ML tools in terms of legitimacy, as the underlying sense-making process is not transparent by design, and the pool of accessed expertise in our study is rather narrow. On the other hand, we found that the process of collectively annotating data provides a fruitful starting point for interdisciplinary exchange and sharing experiences from sense-making activities. Additionally, by creating a labelled data set, we can control input data for the model and guide the learning process.

Nevertheless, how can an open and dynamic assessment of knowledge flows, as supported by ML approaches, be operationalised and implemented? In the following, we highlight relevant key points that guided us through the research process and build the basis for a first reflection of the potential contribution of ML tools to identify evidence for potentially relevant foresight knowledge. We derive the key issues from the leading research questions and the interdisciplinary areas of our expertise (ML, foresight, and science communication research).

Applied ML Approach: Scope and Leading Research Questions

Our study aimed to explore the potential of the chosen ML-supported method for identifying and extracting information that could be used to extend the corpus of data sources that can be used in a foresight process, in particular, for horizon scanning. We wanted to test if ML techniques can help address the pitfalls of foresight when collecting and synthesising the expert and actor's opinions for decision making, i.e., the 'black box' of foresight (Loveridge, 2004, p. 37). The apparent advantage of ML is that it can analyse a much higher number of data sources, thus, potentially broadening the knowledge base for foresight processes. However, we also investigated the question of how ML tools may be used to automate parts of the process and contribute to making knowledge and expert selection processes more transparent.

Our study was led by two main assumptions. First, we wanted to verify if our observation that COVID-19 has led to an increased presence of individual researchers and actors, who are making forward-looking statements or stating trends in mass media publications, can be backed by the data extracted from online news media. To address this, we first searched our corpus data (see Section Corpus Data) for forward-looking statements, and subsequently identified the names of actors mentioned in the context of the extracted statements. The key questions in this context were: Who made the statement or is connected to it? Which contextual information about the actors can we retrieve to verify their (scientific) background?

---

[2]Two internal (2018-2019 and 2019-2020) projects tasked to identify emerging topics with potentially high impact for applied research conducted at Fraunhofer.

TABLE 1 | Data sources and number of extracted news articles.

Publication	Year	Total number of articles	Number of articles mentioning (COVID-19) pandemic
CNA (Channel News Asia)	2019	69,811	12
	2020	64,842	37057
The washington post	2019	50,395	7
	2020	31,361	15,138
CNN	2019	25,204	7
	2020	26,780	13,261
Al jazeera english	2019	15,081	1
	2020	17,369	8,256
Hindustan times	2019	12,745	0
	2020	23,401	10,531
BBC news	2019	9,593	5
	2020	9,107	4,645
The wall street journal	2019	9,311	1
	2020	6,767	3,160
POLITICO.eu	2019	4,685	0
	2020	5,202	2,921
The economist	2019	4,448	3
	2020	3,998	2,131
Total		390,100	97,136

Our second assumption was that by collecting additional information about the publication record of the identified actors, we can verify their (scientific) background. To get additional contextual information about the actors making the statements, we retrieved information about their publication records from the Dimensions[3] database.

Corpus Data

In our empirical study, we searched for potential forward-looking statements from full texts of news articles previously extracted from a wide range of online news portals. The news articles' database that we used was created to cover different world regions and reporting styles, as reported in **Table 1**. The data surveyed in this work is harvested through the use of the commercial NewsAPI, a news interface that can be used to easily query news databases and aggregate the results into a comprehensive data set[4] In total, we collected 447,486 articles published between 1 January 2019 and 31 December 2020.

To retrieve additional information about the background of the actors mentioned in the context of the forward-looking statements, we compared our results with the most recent publication record of the identified persons. We retrieved the data for this comparison through the Dimensions database, which stores information on more than 119 million publications. This includes author names, type of publication (i.e., journal

article, conference proceedings, book chapter, pre-print), date of publication, and the abstract of each publication (full texts are usually not available due to copyrights). We used this information to identify actors with a publication record.

Language and Data Selection Bias

We restricted our corpus data to English-language news articles due to their availability and the volume of articles. The consequence is that the data that we worked with excluded any insights from news articles in other languages, as well as parts of the scientific discourse. However, it can be assumed that in this scenario of a worldwide pandemic, a large part of events and research are discussed and picked up by English-language media and are published in English. Additionally, we tried to compensate for this, as well as to account for regional bias, by selecting news outlets from multiple origins, both geographically and culturally:

- European/US: The Washington Post, Cable News Network (CNN), British Broadcasting Corporation (BBC) News, The Wall Street Journal, POLITICO, The Economist
- Asian: Hindustan Times, Channel News Asia
- Arabic: Al Jazeera English
- African: allAfrica.com

While our approach resulted in a large corpus to analyse, it did not cover the whole media discourse on the COVID-19 pandemic. For example, the sources from Middle- and South America are missing, as well as sources from China. The Global North is represented through most sources (although not the most articles), but, likely, its view is also mirrored in other sources.

Another issue concerns the selection bias of the chosen data sets. Clearly, from the first results, the actors identified from the news media articles are mainly politicians. We strongly assume that the reason for this bias towards politicians results from the chosen dataset of news media. Sources for foresight knowledge typically include foresight reports, trend monitoring reports, recent scientific articles, technology assessment studies, risk reports, and the like. Still, we find it interesting that news media mainly show politicians as the originators of forward-looking statements. For future research, we recommend testing big data sets stemming from different sources and may also further elaborate the adequate and suitable data sources and data sets for identifying foresight knowledge.

Our pragmatic approach of comparing the names extracted with the publication records from the Dimensions' database also led to important limitations. First, although the algorithm can recognise the name(s) in the text snippet that was previously identified as a forward-looking statement, it cannot differentiate between the person mentioned being the source or author of the statement, or the person just being mentioned in the context of the statement. This means that without manually checking the statements extracted, it is not possible to be sure that the number of statements attributed to a specific expert is accurate.

A second issue of the Dimensions database that we used was that in some cases, historical persons were also included as authors. We found out that this is the case when, e.g., a text

---

[3]https://www.dimensions.ai/
[4]https://newsapi.org/

written by the person is published in the context of a modern edition. The examples that we encountered are Abraham Lincoln and Adolf Hitler.

A common issue related to the identification of individuals is the disambiguation of shared names. If two or more individuals share the same name, the algorithm will identify them as being the same person. As we do not have additional information available in our data corpus that could be used for disambiguation (e.g., institutional affiliation), this example can directly translate into wrongly assigning scientific background (or a different background) to one person. In our manual review of extracted statements attributed to the persons with more than two mentions, we compared the statements and the three most recent abstracts from their publication record. Based on that comparison we found that disambiguation is most likely an issue for the following person names from our manually reviewed sample: Sanjay Gupta, Michael Ryan/Mike Ryan, Stephen Hahn, and Rand Paul. Therefore, we did not include them in the list filtering out the persons without expert background. For these cases, the paper abstracts were from completely unrelated fields compared to the statements.

Within the scope of our experimental study, we did not explicitly discuss gender aspects. An interesting question for future research would be to analyse how many male and female actors can we identify in the data. For that, the Worldwide Gender Dictionary[5] could be used.

METHODOLOGICAL APPROACH

Identification and Classification of Forward-Looking Statements

For the work presented here, we created a simple, yet robust classification and enrichment approach that makes use of previous works and freely available data sources. In the following, we detail and describe the major steps of the applied semi-supervised classification approach.

In the first step, we searched the news articles corpus (see Section Corpus Data) for potential forward-looking statements. To create the first data set to work with, we started by defining a loose set of associated keywords that we considered suitable for searching forward-looking statements. Based on this set of keywords indicating future-oriented statements (i.e., 'post-COVID', 'scenario', 'prognosis', 'outlook', 'roadmap', 'impact', 'future', 'future generation', and 'next generation') we automatically searched the data set to collect a set of text snippets, which might contain forward-looking statements through keyword matching (regular expressions). Next to the sentence identified as a forward-looking statement by the algorithm, we also included one sentence before and one after the statement as it proved to be helpful for the manual classification.

In the second step, we reviewed and annotated a random sample of about 560 entries to create a training data set usable for ML classification. The goal was to confirm which of the text parts are forward-looking statements, and

which ones do not contain forward-looking statements. To classify the sample statements, we started from the list of keywords previously used and defined additional attributes indicating forward-looking statements, which are detailed below. Our goal was not to create a generic definition of forward-looking statements, but to develop a preliminary working definition based on which we could classify the random sample statements. Therefore, we manually annotated the dataset and categorised the statements based on considerations of the foresight concept, such as the novelty of the statement or its long-term reference (see Section What is Foresight?), and the properties already identified from the random statements during our review of the sample. Further discussions within the project team, helped refine the working definition of forward-looking statements, along the ML training process.

The working definition of forward-looking statements that we used includes: 1) characteristic elements for forward-looking statements, and 2) elements that we did not consider indicating future-oriented knowledge. The following set of characteristics defining forward-looking statements was the basis for our classification of the sample data set:

- Assumptions, speculations, uncertain/unknown scenarios, or predictions (e.g., *Due to the evolving coronavirus situation, we are facing a period of uncertainty regarding the potential impact on both our supply chain and customer demand.*)
- Long-term prognosis, scenarios, outlooks, and forecasts (e.g., *The economic impact will be devastating on both sides of the border.*)
- Expectations (e.g., *There is also a real danger that more people could potentially die from the economic impact of COVID-19 than from the virus itself.*)
- Warnings (e.g., *Analysts cautioned that a recent rebound in COVID-19 cases could impact activity and consumption.*)
- Estimations, future recommendations (e.g., *It would be enough to offset the impact of tax credits for low- and middle-income taxpayers.*)
- Interviews or survey results including evidence for future impacts, visions, etc. (e.g., *A similar poll found the national economy will grow at its slowest pace in the current quarter since the financial crisis.*)
- Open questions (e.g., *How the virus will impact our economy remains an open question.*)

We agreed to exclude very short-term predictions, such as death rates and weather predictions, from the beginning. Though they are forward-looking in their nature, the added value in terms of new information remains rather limited. However, we tried to include short- and mid-term statements (addressing a time horizon of 1–5 years), when they had a strong foresight character, to include as many example statements as possible in the training data set for the ML algorithm. Also, the short-term forward-looking statements could indicate relevant thematic areas in connection to COVID-19, that might become more relevant in the long term (e.g., future of education, future of work, growing social inequality, impact on suicides, etc.).

---

[5]https://ideas.repec.org/c/wip/eccode/10.html

The following attributes were used to exclude statements that did not meet our requirements for forward-looking statements:

1. Concrete plans, promises, and conclusions for actions with no long-term perspective (e.g., *'According to the panel, the athletes will be permitted to compete provided they've not been subject to suspension.'*)
2. Very short-term predictions, outlooks, and forecasts (e.g., death rates, weather predictions)
3. Retrospective scenarios (e.g., which could have happened, but did not)
4. Interviews or survey results that include evidence for already incurred and observed impacts (view into the past, e.g., *'According to a COVID-19 impact survey, the coronavirus pandemic has robbed 70% of citizens of some form of income.'*)
5. Generalised prospects without concrete scenario or reference (e.g., *'Today's leaders will form our future.'*)

In the third step, we used the resulting manually annotated data (i.e., confirmed forward-looking statements and other statements containing the same keywords) as ground-truth to train and test the ML algorithm. Out of 96,111 statements, which contained the earlier described forward-looking keywords, we identified 42,489 (2019: 27,195; 2020: 15,294) statements as being forward-looking statements according to our definition. In this classification, we did not exclude statements without direct mention or citation of a person as we wanted to train the ML algorithm to recognise forward-looking statements without that restriction. Our sample data included forward-looking statements from various areas, e.g., politics, climate, health, and economy.

Applied Machine Learning Methodology

The manually created ground-truth data set described above was used as input for a (relatively small) deep neural network that is applied to text classification. We provide the resulting ML model on GitHub, along with the rest of the code that we used[6].

To improve the performance of our model, we encoded the input beforehand with the so-called word embeddings; in our case BERT embeddings[7]. What these BERT embeddings do is essentially increase the information depth available from a given text, to emulate how humans read and understand texts. Put briefly, the model gets additional knowledge about each word; for instance, which other words are similar to it; different semantic meanings based on the context; synonyms; etc. In 2018, these embeddings have been described as a major technological breakthrough for the field of natural language processing, see Devlin et al. (2019), allowing researchers to achieve previously unattainable performance on various challenges, including text classification, named entity recognition, text comprehension, machine translation and more. Making use of pre-trained models as in the work presented here has the advantage of being able to use innovative techniques while working with restricted computational resources.

Throughout the training process, we reviewed the initial results, corrected them, and then included those into the training data. We repeated this step twice, and subsequently, reached accuracy scores of 0.98 (0.8 f1-score) on training, 0.75 (0.65 f1) on test, and 0.74 (0.63 f1) on validation data. For the following steps, we only worked with text snippets classified as forward-looking, a total of 27,195 for 2019 and 25,633 for 2020, of which 15,294 come from the articles related to the COVID-19 pandemic.

Next, we used the named entity recognition (NER) from flairNLP, as described by Akbik et al. (2019)[8], to identify persons and organisations from these text snippets. The purpose was to learn who is cited or simply mentioned within a given statement. These NER models are pre-trained in such a way that they achieve high accuracy in extracting the names of persons from texts, without necessarily knowing a specific name beforehand. To prepare a further analysis, we only kept the names of persons that contained at least two elements (first and last name).

Then, we queried the Dimensions database with each name, to identify those who had at least one publication (i.e., paper, book, and chapter) listed in this database (Hook et al., 2018). For this, we searched records of the type 'publications', which, in the list of authors (field 'authors'), contain the name of a person we extracted from our source data, thus, linking that person to a list of publications[9].

RESULTS: ML CLASSIFICATION OF FORWARD-LOOKING STATEMENTS

The results reported below need to be considered the first results of an experimental study, which would certainly benefit from further research and discussion. In Section Discussion, we include our thoughts about the extent to which the combined ML approach can support the following: 1) the identification of forward-looking statements, and 2) provide additional contextual information about the scientific publication record of the actors, which can provide us insights about their contextual knowledge in return.

As briefly described in Section Applied Machine Learning Methodology, the neural network in classifying forward-looking statements worked reasonably well, i.e., we were able to identify such statements from a large text corpus, although, limitations apply, which are discussed in Section Identification and Classification of Forward-Looking Statements. Since this process is not the focus of this work but the foundation for further qualitative and quantitative analysis, we now describe the insights we gained.

First, to get a quantitative understanding of our data, we present in **Table 2** an overview of the number of articles and short statements. The table lists how many of them are classified as forward-looking and how many are from articles mentioning the COVID-19 pandemic. Comparing 2019 and 2020 year-over-year, we observed a small decline in both absolute numbers of articles (6.01%) and the number of forward-looking statements

---

[6]https://github.com/janrn/scicomm-evidence
[7]Specifically, we used https://tfhub.dev/tensorflow/small_bert/bert_en_uncased_L-4_H-768_A-12/2.

[8]https://github.com/flairNLP/flair
[9]See Dimensions API documentation at https://docs.dimensions.ai/dsl/datasource-publications.html.

TABLE 2 | Corpus overview, results of the classification process.

	2019	2020
Total number of articles	201,496	189,391
Number of (potential) forward-looking statements	27,195 (13.49%)	25,633 (13.53%)
Number of articles mentioning Covid-19 pandemic	34 (0.02% of total)	97,447 (51.45% of total)
Number of (potential) forward-looking statements from articles mentioning Covid-19 pandemic	0	15,294 (8.08%)

TABLE 3 | Unique person names from forward-looking statements. In brackets, we give the number of persons that can be found in Dimensions (* compared with total of 2019).

	2019	2020	Overlap between 2020 and 2019
Number of person names	3,405 (1,980)	3,725 (2,872)	618 (533)
Number of person names (from pandemic-related articles)	0	2,262 (1,409)	342 (228)*

(5.74%) extracted from them. However, the majority of those statements in 2020 appear in the articles mentioning the COVID-19 pandemic (59.67%). The overall share of articles mentioning the pandemic is also high at 51.45%.

In the next step, where we further examined the forward-looking statements and extracted names of persons who are cited or otherwise mentioned, we looked up the names in the Dimensions database to identify those who have at least one publication listed in a scientific database. We did not differentiate between the first- or co-authorship or type of publication. The results of this are presented in **Table 3**. For technical reasons, we count only those who appear with two name parts, presumably first and last name. We found that the overall number of unique person names increased slightly from 3,405 to 3,725 (a plus of 9.4%), but the number of those who can be linked to an indexed publication raised (by 45.05%) to a total share of 77.1% in 2020, compared to 58.15% in 2019. Within the subset of pandemic-related articles, this rises to a share of 62.29%. Almost half of the names with a scientific publication entry in 2020 are from news articles linked to the pandemic.

Overall, the quantitative results suggest that we can observe an increase in persons that pass at least a very basic verification of their scientific publication record in the context of the COVID-19 pandemic, coupled with or leading to an increase across all areas of articles. However, the manual check of the identified actors somewhat relativises this result, as explained below.

To get a qualitative understanding of the efficiency of the ML algorithm for identifying the background of the actors mentioned in the statements, we manually reviewed the most frequently

TABLE 4 | Twenty most mentioned names identified in the extracted forward-looking statements. In brackets, we give different spellings for person names, wherever applicable.

Name	Counts	Profession
Donald Trump (Donald J. Trump)	125	politician
Anthony Fauci (Anthony S. Fauci)	36	epidemiologist
Mitch McConnell	28	politician
Jerome H. Powell (Jerome Powell	18	politician
Antonio Guterres	15	politician
Boris Johnson	15	politician
Robert Redfield	14	virologist
Barack Obama	12	politician
George Floyd	11	citizen
Nancy Pelosi	11	politician
Xi Jinping	10	politician
Kamala Harris	10	politician
Tedros Adhanom Ghebreyesus	10	epidemiologist, WHO head
George W. Bush	10	politician
Bernie Sanders	9	politician
Hillary Clinton	9	politician
Tom Frieden	7	physician

mentioned actors in the statements extracted by the algorithm. We compared the statements and the abstracts/titles of the most recent three publications of the identified actors. **Table 4** shows the 20 most mentioned names in the extracted statements, which include, next to renowned epidemiologists and virologists, a large number of politicians. This result is not surprising, as Dimensions is not limited to scientific literature, and a few politicians published scientific work, albeit, in some cases many years ago. Therefore, the quantitative output of this research step needs to be complemented by manual review and additional research to achieve an understanding of the background of the actors mentioned.

Thus, in an ulterior manual step, we did online research to manually filter out the individuals without a scientific background. **Table 5** shows the resulting list of most cited actors with a scientific background in the context of COVID-19 from our analysed corpus data. Not surprisingly, the most cited actors in the news items identified by the ML algorithm correspond to famous COVID-19 advisors from major institutions in the English-speaking world, e.g., Anthony Fauci (NIH), Robert Redfield (CDC), and Tedros Adhanom Ghebreyesus (WHO).

By comparing the frequency of the actors mentioned in the extracted statements, we could clearly see a trend where actors were cited the most. Also, we compared the frequency of actors cited with the previous year (see **Table 3**): in 2020, 618 names from 2019 can be identified again, 342 of those appeared in texts related to the COVID-19 pandemic.

DISCUSSION

In this chapter, we discuss the opportunities, challenges, and limitations of our approach. We summarise our thoughts and

TABLE 5 | Most cited actors in context of COVID-19 after manually filtering out the names of individuals without scientific background.

Name	Count	Expertise
Anthony (S.) Fauci	36	epidemiologist
Robert Redfield	14	virologist
Tedros Adhanom Ghebreyesus	10	epidemiologist, WHO head
Tom Frieden	7	physician
Nancy Messonnier	6	physician
Deborah Birx	5	physician, diplomat
Richard Besser	3	doctor, executive
Michiko Ueda	3	political scientist
Christopher Murray	3	researcher
Ramanan Laxminarayan	3	economist, epidemiologist
Leana Wen	3	physician, columnist, former health commissioner

conclusions with regards to the potential that our approach can have for sense-making processes in foresight.

Relevance of ML for Horizon Scanning: Insights and Limitations

Our results suggest that the extracted data can be useful for answering the following questions: Which forward-looking statements can be identified in contemporary news articles? Who is making the statement? Which are the latest publications of the person making the statement?

In our view, our ML-aided approach is able to provide relevant insights in the context of forward-looking statements in public media. Our results confirmed that the ML approach is a helpful support for semi-automatically extracting forward-looking statements from a large dataset and for identifying actors associated with those.

The four examples in **Table 6** of statements extracted by the algorithm show that the identification of forward-looking statements with the COVID-19 connection worked reasonably well for many cases. However, the long-term orientation and novelty of the identified forward-looking statements remain unsatisfactory in large parts. An example is the statement attributed to Beth Tarasawa, which was positively recognised as a forward-looking statement by the algorithm, but only with a rather low score of 0.68 (*[They] predict that extended school closures could potentially cause serious academic setbacks for students struggling to adapt to remote instruction.*). According to our definition, the forward-looking statement should have received a higher score as it associates an emerging issue with COVID-19, which could be relevant for a specific foresight activity. In comparison, a statement by Annegret Kramp-Karrenbauer was identified as a forward-looking statement by the algorithm with a very high score of 0.94. However, the statement is not a forward-looking statement according to our definition (*Annegret Kramp-Karrenbauer unexpectedly*

announced that she will not run for chancellor in the 2021 general election.). Although this inaccuracy of results is predictable due to the fuzzy concept of forward-looking statements that we used, the resulting inaccuracy has implications for the fitness for purpose of the extracted statements in a foresight context. As noted in 3.1, our working definition for a forward-looking statement included short- and medium-term elements to have as many examples of the statements in the ground-truth data. This means that the statements identified by the ML algorithm do not necessarily include only the long-term perspective, which is important in the context of adequate and appropriate foresight.

Also, the comparison with the publication record of the identified actors provided meaningful contextual information that can add relevant information about the (scientific) background of the actors.

In summary, politicians were identified in most of the extracted statements. Also, we discovered that many statements were made by scientists and practitioners (see **Table 4**). However, our simplistic approach to base the identification and classification of actors solely on records of recent publications is limited. The problem is that the publication record of a person does not necessarily reflect the various roles that he or she has. An example is Tedros Adhanom Ghebreyesus, who resulted in the third most mentioned actor with a scientific background in the context of the extracted COVID-19 statements (see **Table 5**). His scientific publication record in Dimensions confirms his expertise as an epidemiologist. However, he is also the Director-General of the WHO, which, on the one hand, has an influence on his visibility in the media. On the other hand, the data that we analysed do not give us the possibility to assess whether his statements were made based on his scientific expertise or in his function as WHO director. Another related example is Rand Paul, who is a physician but also a politician. This additional level of complexity needs to be critically taken into account for the manual selection of data as well as the interpretation of the results from the ML analysis.

Within the foresight process, politicians are usually the target audience for whom foresight knowledge is generated (e.g., recommendations based on future-robust knowledge). To integrate their forward-looking expertise is even more difficult, as it might be influenced by political agendas and could, therefore, undermine the scientific knowledge base or might even add an unwanted normative framing to the knowledge generation process.

Added Value for Sense-Making Processes

Clearly, the ML approach is limited in terms of automatically identifying tangible evidence. It rather supports the identification of signals of the potentially relevant evidence in massive amounts of data, which again requires supervised examination to make sense of the results.

We want to emphasise that an in-depth discussion and assessment of the quality of the collected statements were outside the scope of this experimental study. We focused our analysis on testing the applicability of the ML approach for the extraction of relevant information sources from a large data set of media content that can ultimately contribute to strategic foresight

TABLE 6 | Exemplary statements extracted by the algorithm and prediction score.

Person name	Identified forward-looking statement used for comparison	Forward-looking statement prediction score
Anthony Fauci	The Food and Drug Administration could authorize the vaccines for emergency use by mid- to late December. You don't want to get ahead of yourself and claim any victories, but this has the makings of a very, very important positive impact on ending this outbreak, said Anthony S. Fauci, director of the National Institute of Allergy and Infectious Diseases. A vaccine that is highly efficacious, if taken by a very, very high percentage of people, could potentially crush this outbreak similar to what was done with outbreaks of measles and polio and smallpox and other diseases.	0.99
Anthony Fauci	Dr. Anthony Fauci, infectious diseases chief at the National Institutes of Health, in several television interviews on Friday, said more tests would be available over the next week, but that officials should not wait before trying to mitigate the virus's effects. "We will have considerably more testing in the future, but you don't wait for testing," Fauci said on "CBS This Morning." He said school closing and similar measures are "generally an appropriate approach".	0.99
Beth Tarasawa	Using existing data on learning loss typically seen in the summer from a national sample of over five million students in grades three through eight, Kuhfeld and her colleague, Beth Tarasawa, predict that extended school closures could potentially cause serious academic setbacks for students struggling to adapt to remote instruction. In a worst-case scenario, they may retain only 70% of the gains they had made in reading and only 50% of the gains made in math. Math could be a particular sticking point, Kuhfeld said.	0.68
Annegret Kramp-Karrenbauer	While some of the budget provisions could pass the Democratic House, others, like the elimination of the student loan forgiveness plan, will have a much harder time getting through. Germany's political future is mired in uncertainty after longtime Chancellor Angela Merkel's hand-picked successor, Annegret Kramp-Karrenbauer, unexpectedly announced that she will not run for chancellor in the 2021 general election. Merkel, who has led Germany for 15 years, is not planning to run for reelection.	0.94

processes. Although we see the potential for our approach to providing data that can contribute to assessing the quality of the extracted statements, this step requires further research and thorough methodological discussions. As already highlighted, the data identified and extracted using the tested ML algorithm mainly allowed us to recognise preliminary insights. Although the insights from the extracted contextual information of news texts can be used to determine the (scientific) background of the actors to a certain extent, our experience showed that the extracted data needs a thorough qualitative review to be used as evidence-based for further interpretations.

One of our learnings from the interdisciplinary approach in combining ML, foresight, and media analysis of science communication methodologies is that a close combination of qualitative and quantitative methods, as well as interdisciplinary exchange and interaction, are key in all stages of an ML-supported research process. Starting from the decision about which data sets to be used for the analysis, over the discussion and review of keywords and key concepts guiding the ML model training, as well as data set classifications; all these stages profit from the interdisciplinary discourse and reflection. The interdisciplinary setting forced us to think out of the box and to re-assess and clarify key concepts needed for the ML classification. Also, the discussion and interpretation of results considerably profit from the input from various disciplines and models of thought.

What we realised during this process is that experts are needed for all phases of the ML-aided process. However, it also made it clear that building a reliable ML model, which can handle dynamic and practice-related concepts, is a time-consuming endeavour that needs considerable work compared to what we were able to do within the scope of this study. Setting up such an ML model is more challenging compared to targeted research conducted by a person. Initial costs and efforts needed to create such a system are high and need to be taken into consideration (Mauksch et al., 2020, p. 10).

Outlook and Future Research

Enhancing the transferability of accessible evidence is key to fostering collaborative research between researchers and practitioners. We think that an ML approach with a strong interdisciplinary focus has the potential to enhance the transferability of evidence between various disciplines and practices. The procedural setting of an ML approach and the requirement of ongoing refinement and further training of ML models show parallels with sense-making processes. The data that is made accessible during an ML annotation and classification process build common ground for reflecting on implicit knowledge and expertise. These seem to be suitable features to build 'enquiry machines (EMs)[…], interactive machines to explore sociocultural topics' (Jungnickel, 2020, 4, p. 36).

Interdisciplinary and collaborative approaches have been proven to generate opportunities for creating and transferring new knowledge, even beyond the academic borders (Coulter, 2013). The potential of the ML tools to

help open a research process to a wider audience and to contribute to more transparent sense-making processes is high in our view. Especially, when considering the idea of creating 'machines for making and communicating research' (Jungnickel, 2020), the ML tools could provide a useful basis for exchange and discussion. This consideration that relates to the societal dimension of knowledge production and science communication could be further explored (Tuebke et al., 2001; Gibbons et al., 2010). In our view, the endeavour to identify evidence in forward-looking statements from public news media data raises questions that are highly relevant for the relationship between science and society. Here, we see a strong potential for the application of similar ML-supported approaches in the field of science communication research.

DATA AVAILABILITY STATEMENT

The data analysed in this study was obtained from Dimensions, the following licenses/restrictions apply: Requests to access the datasets must first be approved by the Dimensions Review Committee. Requests to access these datasets should be directed to https://dimensions.freshdesk.com/support/solutions/articles/23000020558-how-can-i-get-access-to-dimensions-for-scientometric-research.

AUTHOR CONTRIBUTIONS

All authors contributed equally. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Ahmed, N., Shahbaz, T., Shamim, A., Shafiq Khan, K., Hussain, S. M., and Usman, A. (2020). The COVID-19 infodemic: a quantitative analysis through Facebook. *Cureus* 12, e11346. doi: 10.7759/cureus.11346

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019). "FLAIR: an easy-to-use framework for state-of-the-art NLP," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (Minneapolis, MN: Association for Computational Linguistics), 54–59.

Chipidza, W., Akbaripourdibazar, E., Gwanzura, T., and Gatto, N. M. (2021). Topic analysis of traditional and social media news coverage of the early COVID-19 pandemic and implications for public health communication. *Disaster Med. Public Health Prep.* doi: 10.1017/dmp.2021.65. [Epub ahead of print].

Coulter, J. (2013). "Interdisciplinarity: creativity in collaborative research approaches to enhance knowledge transfer," in *Innovation through Knowledge Transfer 2012. Smart Innovation, Systems and Technologies*, eds R. Howlett, B. Gabrys, K. Musial-Gabrys, and J. Roach (Berlin; Heidelberg: Springer).

Cuhls, K., Erdmann, L., Warnke, P., Toivanen, H., Toivanen, M., Van der Giessen, A., et al. (2015). *Models of Horizon Scanning - How to integrate Horizon Scanning into European Research and Innovation Policies*. Brussels: European Commission.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1* (Minneapolis, MN: Association for Computational Linguistics), 4171–4186.

European Commission (2020). 2020 *Strategic Foresight Report – Charting the course towards a more resilient Europe*. Available online at: https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/strategic-foresight/2020-strategic-foresight-report_en.

Gavigan, J. P., Scapolo, F., Keenan, M., Miles, I., Farhi, F., Lecoq, D., et al. (2001). "A practical guide to regional foresight," in *European Commission Research Directorate General, STRATA Programme, ed FOREN Network* (Seville: Institute for Prospective Technological Studies).

Georghiou, L., Harper, J. C., Keenan, M., Miles, I., and Popper, M. (2008). *The Handbook of Technology Foresight: Concepts and Practice*. PRIME Series on Research and Innovation Policy: Edward Elgar.

Geurts, A., Gutknecht, R., Warnke, P., Goetheer, A., Schirrmeister, E., Bakker, B., et al. (2021). New perspectives for data-supported foresight: the hybrid AI-expert approach. *Fut. Foresight Sci.* e99. doi: 10.1002/ffo2.99

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., and Trow, M. (2010). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: SAGE Publications Ltd.

Hook, D. W., Porter, S. J., and Herzog, C. (2018). Dimensions: building context for search and evaluation. *Front. Res. Metr. Anal.* 3, 23. doi: 10.3389/frma.2018.00023

Ioannidis, J. P. A., Salholz-Hillel, M., Boyack, K. W., and Baas, J. (2021). The rapid, massive growth of COVID-19 authors in the scientific literature. *R. Soc. Open Sci.* 8, 210389. doi: 10.1098/rsos.210389

Jasanoff, S., and Kim, S.-H. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press, London.

Jungnickel, K. (2020). *Transmissions: Critical Tactics for Making and Communicating Research*. Cambridge, MA: The MIT Press.

Krawczyk, K., Chelkowski, T., Laydon, D. J., Mishra, S., Xifara, D., Gibert, B., et al. (2021). Quantifying online news media coverage of the COVID-19 pandemic: text mining study and resource. *J. Med. Internet Res.* 23, e31544. doi: 10.2196/31544

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Loveridge, D. (2004). Experts and foresight: review and experience. *Int. J. Foresight Innovation Policy* 1, 33–69. doi: 10.1504/IJFIP.2004.004651

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

Mauksch, S., Von der Gracht, H. A., and Gordon, T. J. (2020). Who is an expert for foresight? A review of identification methods. *Technol. Forecast. Soc. Change* 154, 119982. doi: 10.1016/j.techfore.2020.119982

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. AI Magazine.

Miller, R. (2018). *Transforming the future: anticipation in the 21st century*. Paris; Abingdon, UK; New York, NY: UNESCO; Routledge.

Muller, M., Guha, S., Baumer, E. P. S., Mimno, D., and Shami, N. S. (2016). "Machine learning and grounded theory method: convergence, divergence, and combination," in *Proceedings of the 19th International Conference on Supporting Group Work (GROUP '16)* (New York, NY: Association for Computing Machinery), 3–8.

O'Connor, C., O'Connell, N., Burke, E., Nolan, A., Dempster, M., Graham, C. D., et al. (2021). Media representations of science during the first wave of the COVID-19 pandemic: a qualitative analysis of news and social media on the island of Ireland. *Int. J. Environ. Res. Public Health* 18, 9542. doi: 10.3390/ijerph18189542

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Why should i trust you?: explaining the predictions of any classifier. *arXiv:1602.04938*. doi: 10.1145/2939672.2939778

Robinson, D. K. R., Schoen, A., Larédo, P., Molas Gallart, J., Warnke, P., Kuhlmann, S., et al. (2021). Policy lensing of future-oriented strategic intelligence: an experiment connecting foresight with decision making contexts. *Technol. Forecast. Soc. Change* 169, 120803. doi: 10.1016/j.techfore.2021.120803

Rosa, A. B., Kimpeler, S., Schirrmeister, E., and Warnke, P. (2021). Participatory foresight and reflexive innovation: setting policy goals and developing strategies in a bottom-up, mission-oriented, sustainable way. *Eur. J. Futures Res.* 9, 2. doi: 10.1186/s40309-021-00171-6

Tuebke, A., Ducatel, K., Gavigan, J., and Moncada-Paternò-Castello, P. (2001). *Strategic Policy Intelligence: Current Trends, the State of Play and Perspectives*, ed JRC-IPTS. Seville: European Commission.

Von Schomberg, R., Guimaraes Pereira, A., and Funtowicz, S. (2006). "Deliberating foresight knowledge for policy and foresight knowledge assessment," in *Interfaces Between Science and Society*, eds S. Guedes Vaz, A. Guimarães Pereira, and S. Tognetti (Sheffield: Greenleaf Publishing).

Warnke, P., and Schirrmeister, E. (2016). Small seeds for grand challenges— Exploring disregarded seeds of change in a foresight process for RTI policy. *Futures* 77, 1–10. doi: 10.1016/j.futures.2015.12.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.