Article

A Holistic Framework for Forecasting Transformative AI

Ross Gruetzemacher

Systems and Technology, Auburn University, Auburn, AL 36849, USA; rossg@auburn.edu

Received: 1 June 2019; Accepted: 21 June 2019; Published: 26 June 2019

Abstract: In this paper we describe a holistic AI forecasting framework which draws on a broad body of literature from disciplines such as forecasting, technological forecasting, futures studies and scenario planning. A review of this literature leads us to propose a new class of scenario planning techniques that we call scenario mapping techniques. These techniques include scenario network mapping, cognitive maps and fuzzy cognitive maps, as well as a new method we propose that we refer to as judgmental distillation mapping. This proposed technique is based on scenario mapping and judgmental forecasting techniques, and is intended to integrate a wide variety of forecasts into a technological map with probabilistic timelines. Judgmental distillation mapping is the centerpiece of the holistic forecasting framework in which it is used to inform a strategic planning process as well as for informing future iterations of the forecasting process. Together, the framework and new technique form a holistic rethinking of how we forecast AI. We also include a discussion of the strengths and weaknesses of the framework, its implications for practice and its implications on research priorities for AI forecasting researchers.

Keywords: AI forecasting; technology forecasting; scenario analysis; scenario mapping; transformative AI; scenario network mapping; judgmental distillation mapping; holistic forecasting framework

1. Introduction

In a world of quick and dramatic change, forecasting future events is challenging. If this were not the case then meteorologists would be out of a job. However, meteorological forecasting is relatively straightforward today given the relatively low price of computation, the advanced capabilities of numerical simulation and the myriad powerful sensors distributed around the world for collecting input information. Forecasting technological progress and innovation, however, is much more difficult because there is no past data to draw upon and future technologies are at best poorly understood [1]. Forecasting progress toward broadly capable AI systems is even more difficult still because we do not yet know the fundamental architectures that may drive such systems.

This decade has seen significant milestones in AI research realized [2–5], and the realization of these milestones has left many to perceive the rate of AI progress to be increasing. This perceived increase in the rate of progress has been accompanied by substantial increases in investment, as well as increased public and governmental interest. Consequently, there is a growing group in the AI strategy research community that is working to measure progress and develop timelines for AI, with significant effort focusing on forecasting transformative AI or human-level artificial intelligence (HLAI). (We consider forecasts for human-level machine intelligence, high-level machine intelligence and artificial general intelligence to be equivalent to forecasting HLAI.). Efforts to these ends, however, are not unified and the study of AI forecasting more broadly does not appear to be directed at a well understood objective. Only one previous study has proposed a framework for forecasting or modeling

AI progress [6]. This paper outlines an alternative to that previous framework that utilizes both judgmental, statistical and data driven forecasting techniques as well as scenario analysis techniques.

To be certain, efforts to forecast AI progress are of paramount importance. HLAI has the potential to transform society in ways that are difficult to anticipate [7]. Not only are its impacts difficult to imagine, but the notion of HLAI itself is ill-defined; what may be indicative of human-level intelligence to some may not be sufficient to others, and there is no definitive test for human-level intelligence. (Recently, a new field of scientific study has been proposed for better understanding machine behavior [8].) This has lead studies concerned with forecasting AI progress or HLAI to focus on the replacement of humans at jobs or tasks [9,10]. The lack of an objective definition for HLAI is due in part to the fact that we do not know how to create it. In theory, HLAI could be instantiated by one algorithm [11] or constructed by combining different components [12]. To adequately address this and other unique challenges faced in forecasting HLAI, methods that integrate diverse information and a variety of possible paths are required.

The necessity of planning for HLAI is obvious. It is also plausible, and perhaps even likely, that AI will have severe transformative effects on society without reaching human-level intelligence. A formal description of the extreme case for such a scenario is Drexler's notion of comprehensive AI services (CAIS) [13]. Therefore, for the purpose of ensuring that AI is developed to do the most good possible for humanity, we identify the primary task of AI forecasting to be that of forecasting transformative AI (this includes artificial general intelligence [AGI], AI generating algorithms and superintelligence). We define transformative AI to be any set of AI technologies that has the potential to transform society in ways that dramatically reshape social structures or the quality of life for social groups.

Here, we take the position that AI forecasts solely in the form of timelines (dates given by which we should expect to have developed transformative AI) are undesirable. To address this issue we propose a new AI forecasting framework along with a new scenario mapping technique that supports the framework. Independently, the framework and the new method each constitute novel contributions to the body of knowledge. However, together the framework and new technique demonstrate a holistic rethinking of how we forecast AI. It is this new perspective that we believe to be the paper's most significant contribution.

In the following pages the paper proceeds by first examining related literature. We do not consider the broader body of literature for the relevant topics, rather we focus only on the salient elements. After outlining scenario planning techniques, we move to propose a new subclass of scenario mapping techniques. Next, we propose a new method as part of this subclass which we call judgmental distillation mapping. This new method is then described as a critical component of the new AI forecasting framework. Following this description of the framework, we discuss strengths, weaknesses, the implications of practice and the implications on future research in AI forecasting. We conclude by summarizing the key ideas and recommendations.

2. Literature Review

This section examines several bodies of literature relevant to the holistic framework being proposed. This literature review is by no means comprehensive, and, due to the large number of academic disciplines and techniques covered, a more extensive literature is suggested for future work. We consider the research topics of forecasting, technology forecasting, scenario analysis, AI forecasting as well as a brief discussion of digital platforms.

2.1. Forecasting

Forecasting techniques are commonly broken down into two broad classes: judgmental methods and statistical methods [14]. Statistical methods are preferred for most forecasting applications and can range from simple extrapolations to complex neural network models or econometric systems of simultaneous equations [15]. However, statistical methods perform poorly in cases with little or no historical data, cases with a large degree of uncertainty and cases involving complex systems [16].

In such situations it is common to fall back on judgmental techniques. In this subsection we will forgo any discussion of statistical methods to focus on the different judgmental techniques and the challenges of expert selection.

Surveys are likely the most widely used judgmental technique. They solicit expert opinion from multiple experts without interaction between them. This technique is widely used because it is straightforward to implement and relatively inexpensive [1]. Challenges to this method include sampling difficulties, especially those due to nonresponses. The Cooke method (or the classic method) of assessing the quality of expert judgements for expert elicitation comes from the field of risk analysis [17]. It is a very powerful technique that involves the inclusion of calibration questions to calibrate the experts' forecasts so that they may be weighted during aggregation [18].

The Delphi technique was developed at the Rand Corporation in the 1950s at the same time as the development of scenario planning methods [19]. This approach involves a group of experts participating in an anonymized forecasting process through two or more rounds [1]. Each round involves answering questionnaires, aggregating the data and exchanging the summarized results and comments. Expert participation, expert selection and the falloff rate of participants over iterative survey rounds are the primary challenges. The Delphi technique is powerful and versatile, with the capability to be used for scenario building exercises as well as forecasting, and the flexibility to support large groups of experts with small modifications [20]. Despite its wide use for over a half century, there are still many questions about fundamental issues of its effectiveness for certain situations [21,22]. Specifically, academic work that has been conducted on the Delphi technique has frequently used students, which, for numerous reasons may be misleading. No work on the Delphi technique or other powerful judgmental forecasting techniques has been conducted to assess the quality of forecasts for the purpose of technology forecasting.

Prediction markets are exchange traded markets intended for predicting the outcomes of events. They rely on a platform that allows people to make trades depending on their assessment of these outcomes. Prediction market contracts are binary options that are created to represent a forecasting target and then traded through the market. During trading, the market price of a contract adjusts dynamically to account for participants' predictions and is used as an indicator of the probability of these events. This incentivizes participants to be as accurate as possible in order to receive the most gain while allowing for aggregation over an arbitrarily large market. The free market tends to collect and aggregate predictive information well due to the strong economic incentives for better information. Consequently, prediction markets often produce forecasts that have lower prediction error than conventional forecasting techniques [23]. Green et al. performed a comparison of the Delphi technique and prediction markets, finding that, when feasible, prediction markets have some advantages, but that the Delphi technique was still generally underused (it is unclear which is better for technology forecasting) [24]. However, the advantages of each technique were also dependent on the problem. Prediction markets performed better for short-term, straightforward problems whereas the Delphi technique was useful for a broader range of problems and for high uncertainty situations.

Superforecasting is a recently developed technique that uses groups of forecasting experts, i.e., superforecasters, in combination with advanced aggregation techniques to generate forecasts. Superforecasting has been demonstrated to be more accurate than prediction markets and to forecast certain types of targets (e.g., geopolitical events) better than any other methods [25]. The technique was developed using forecasting tournaments for a competition for the US' Intelligence Advanced Research Projects Activity (IARPA). The project was funded for the purpose of developing new methods in order to improve the US intelligence communities forecasting abilities [26]. However, superforecasting is not suitable for all forecasting problems. Particularly, it is ill-suited for predictions that are either entirely straightforward and well suited for econometric methods, or for predictions that are seemingly impossible. It is also not suitable for existential risk applications [27]. Furthermore, while it may be one of the most powerful forecasting methods available for near-term forecasts, it still is not able to

make forecasts any better than a coin toss for events over five years in the future (Tetlock considers experts no better than normal people at forecasting political events).

Combining different types of forecasts that draw from different information sources can be a powerful technique for forecasting when there is significant uncertainty about the situation or uncertainty about the different methods [14]. Another powerful technique can be the adjustment of statistical forecasts using expert judgment, particularly in cases of high uncertainty where domain expertise is critical and environments are poorly defined [28]. In such cases, structured judgmental adjustment can be a very powerful technique as long as efforts are made to counter cognitive biases [29].

Scenario planning methods are sometimes considered an adjunct forecasting method and scenarios are commonly employed to deliver the results of forecasts to decision makers [14]. However, substantial work has been conducted considering their practical use in improving decision making under uncertainty [30,31]. They are considered an essential technique in the technology forecasting and management literature [1], thus, we devote an entire subsection to them in the following pages.

Goodwin and Wright examine both statistical and judgmental forecasting methods in their ability to aid the anticipation of rare, high-impact events [21]. They find that while all methods have limitations, it is possible to combine dialectical inquiry and components of devil's advocacy with the Delphi technique and scenario planning techniques to improve the anticipation of rare events. In their comparison of techniques (including an informative table comparing methods for anticipating rare events) they consider several judgmental methods including expert judgment, structured judgmental decomposition, structured analogies, judgmental adjustment and prediction markets, as well as the Delphi technique and scenario planning.

Selecting experts is a challenging but necessary task when using any of these judgmental forecasting techniques. The first step in identifying experts is to identify the range of perspectives that will be needed in the study [1]. Researchers typically want to prioritize the most knowledgeable experts for vital perspectives first; less vital viewpoints can often times use less knowledgeable experts or substitute secondary sources for expert opinion. Researchers should also be cognizant of possible sources of experts' biases when selecting experts and analyzing their responses. Some significant attributes include a broad perspective relating to their knowledge of the innovation of interest, a cognitive agility for being able to extrapolate from their knowledge to satisfy future possibilities, and uncertainties and a strong imagination [32]. There is also the question of how many experts one needs for a study. This is commonly dependent on many factors, including the type of the study, the technology of interest and the scope of the study. Sampling diverse populations can lead to many issues, however, when it is necessary, documentation for the particular type of study commonly addresses these issues [33].

2.2. Technology Forecasting

Technology forecasting is a challenging task and the body of literature concerning this topic is very broad. To be certain, there is not a well-developed field of study that directly concerns the forecasting of future technologies. Much of what is considered here as technology forecasting literature is focused on technology management, and, consequently, many of the techniques are intended to aid in organizational management and planning.

A wide variety of methods are used for technology forecasting, including both statistical and judgmental techniques. Other techniques are also used, some of which are unique to technology forecasting. Innovation forecasting techniques can be used for mapping scientific domains that rely on bibliometric analysis [34]. Tech mining is a similar technique that harnesses data mining methods to extract information from patent databases and the Internet for the purposes of innovation forecasting [35]. Due to the substantial uncertainty, scenario analysis techniques are also widely used for strategic planning involving emerging technologies [1]. This subsection does not revisit judgmental forecasting techniques discussed in the previous subsection, but focuses on techniques that have not yet been discussed. Scenario analysis is discussed in depth in the following subsection.

Assessing progress—particularly the rate of progress—is essential when developing any type of technology forecasting model. This is so because the naïve assumption that historical trends can be extrapolated to the future is many times correct, and, consequently, trend extrapolation is a very powerful forecasting technique [1]. Indicators are variables that can be used for extrapolation or for building statistical forecasting models because we believe them to be good for predicting future progress. There are two basic types of indicators that are of interest in technology forecasting: science and technology indicators and social indicators. Science and technology indicators, or simply technology indicators, are directly related to the progress of the technology of interest. Social indicators are intended to collectively represent the state of a society or some subset of it. Technology indicators ideally must adhere to three restrictions: (1) the indicator must measure the level of a technology's functionality, (2) the indicator must be applicable to both the new technology and to any older technologies it replaces and (3) there must be a sufficient amount of data available to compute historical values. In reality, many times indicators are not available which satisfy all of these requirements. In such cases efforts should be made to identify indicators that suffice as best as possible. Social indicators can include economic factors, demographic factors, educational factors, etc., and they are thought to be analogous to a technology's functional capacity.

Technology roadmapping is a widely used and flexible technique that is commonly used for strategic and long-term planning [36]. It is known to be particularly effective in structuring and streamlining the research and development process for organizations [37], but it can be used for planning at both an organizational level and a multi-organizational level. It is generally thought to consist of three distinct phases—a preliminary phase, a roadmap construction phase and a follow-up phase—and commonly uses workshops for the map generation phase [38]. When applied, it often uses a structured and graphical technique that enables exploring and communicating future scenarios. However, its lack of rigor and heavy reliance on visual aids can also be seen as weaknesses [1].

Innovation forecasting is a term that is typically associated with the use technology forecasting methods in combination with bibliometric analysis [34]. In general, bibliometric methods are powerful analysis tools for understanding the progression of science. Such methods have been used for the mapping of this progression in different scientific disciplines for several decades [39]. Maps of relational structures present in bibliometric data are useful for visualizing the state of research within the domain(s) of interest and can lead to insights regarding future research directions and geopolitical issues [40].

Tech mining is another notion that is frequently associated with innovation forecasting and management [35]. It generally refers to a broad set of techniques which can be used to generate indicators from data. Porter and Cunningham discuss the use of innovation indicators for understanding emerging technologies, and propose nearly 200 such indicators. Here, we consider tech forecasting to encompass all bibliometric and scientometric techniques used for the purposes of technology forecasting.

While we have focused here on judgmental forecasting techniques and other techniques for technology forecasting, there is evidence that suggests that extrapolation and statistical methods are better for forecasting technological progress [41]. Studies have found that technology forecasts developed using statistical methods were more accurate than those developed from other methods, with forecasts about autonomous systems and computers being the most predictable [42]. However, there is certainly not agreement on this topic. Brynjolfsson and Mitchell conclude that “simply extrapolating past trends will be misleading, and a new framework is needed,” [43]. The holistic perspective proposed here attempts to provide a new framework.

2.3. Scenario Analysis

Scenario analysis is a term used in technology management literature to refer to scenario planning techniques when applied in the context of technology and innovation forecasting [1]. People use scenario analysis naturally by thinking in terms of future scenarios when making most decisions involving uncertainty in everyday life. It is also a very effective technique for decision-making processes in more complex situations [44]. Scenario methods are rooted in strategic planning exercises from the

military in the form of ‘war game’ simulations, or simply wargames. Wargames are a type of strategy game that have both amateur and professional uses. For amateurs they are used for entertainment, with some of the earliest examples being the games of Go and chess. Fantasy role-play games such as Dungeons and Dragons are also derived from wargames and used for entertainment. Professionally, wargames can be used as a training exercise or for research into plausible scenarios for highly uncertain environments such as those encountered on battlefields during wartime [45]. Events in World War II, such as the allied preparations for D-Day, made clear to military commanders the value of wargames and scenario techniques. Following the war, during the 1950s and 1960s, new scenario techniques were independently developed in both the United States and France. In the United States, the methods were developed at the Rand Corporation, a research and development venture of the US Air Force. In France, the techniques were developed for public planning purposes. Although developed independently, these two schools eventually led to the development of very similar scenario techniques.

Scenario analysis, as it is known today, typically involves the development of several different scenarios of plausible futures. It is most widely thought of as a qualitative technique for the purposes of strategic planning in organizations [46]. Proponents of this thinking often consider scenarios as an aid for thinking about the future, not for predicting it. However, a rich body of literature has developed over the years, and many quantitative and hybrid techniques have also been shown to be practically useful [20]. Here we describe three schools of scenario techniques: the intuitive logics school, the probabilistic modified trends (PMT) school and La Prospective, a.k.a. the French school. We attempt to outline these different schools below.

The most prominent of qualitative methods, having received the most attention in the scenario planning literature, is the intuitive logics school [20]. After being developed by at the Rand Corporation in the 1950s and 1960s, it was popularized from its use by Royal Dutch Shell in the 1970s, and it is sometimes referred to as the ‘Shell approach’ [19]. This school of methods is founded on the assumption that business decisions rely on a complex web of relationships including economic, technological, political, social and resource-related factors. Here, scenarios are hypothetical series of events that serve to focus attention on decision-points and causal processes. While such scenario planning techniques are very useful for business purposes, alternative scenario planning techniques can be used for much more than investigating blind spots in organizations’ strategic plans [47].

The most common of quantitative methods is considered to be the PMT school, which also originated at the Rand Corporation during the 1960s [20,48]. This school incorporates two distinct methodologies: trend-impact analysis (TIA) and cross-impact analysis (CIA) [19]. TIA is a relatively simple concept which involves the modification of extrapolations from historical trends in four relatively simple steps. CIA attempts to measure changes in the probability of the occurrence of events which could cause deviations from extrapolated trends through cross-impact calculations. The primary difference between the two techniques is the added layer of complexity introduced in CIA during the cross-impact calculation.

The two schools described above may do well to illustrate qualitative and quantitative scenario techniques, but they are by no means an exhaustive description of this dichotomy of scenario planning methods. Another way to think of qualitative and quantitative scenarios is as storylines and models. The former captures possible futures in words, narratives and stories while the latter captures possible futures in numbers and rules of systems’ behaviors. Schoemaker notably suggests that the development of quantitative models is an auxiliary option for assisting in making decisions, whereas the development of scenarios is the purpose of the activity [49]. Hybrid scenario techniques attempt to bridge the gap between methods which rely on storylines and models.

La Prospective is a school of hybrid scenario techniques that emerged in the 1950s in France for long-term planning and to provide a guiding vision for policy makers and the nation [20]. This school is unique in that it uses a more integrated approach through a blend of systems analysis tools and procedures, including morphological analysis and several computer-aided tools [19]. Although it arose independently, this school can also be seen to a large extent to combine the intuitive logics and PMT

methodologies. A full review of scenario planning literature is beyond the scope of this work, but we believe that these simple characterizations to be sufficient for the purpose of this work.

2.3.1. Scenario Analysis for Mapping

Traditional qualitative scenario planning techniques certainly have a role in assisting decision makers of organizations and other stakeholders involved in the development and governance of transformative AI. However, such techniques can do little to map the plausible paths of AI technology development due to the large space of possible paths. Traditional quantitative methods certainly have a role in some organizational decisions as well. However, while they are commonly sufficient for strategic decision making, they typically fall short for understanding and informing design decisions of complex systems.

Over the past two decades, the use of scenario analysis techniques for mapping complex systems, complex environments and complex technologies has increased [50]. Particularly, we focus on three such techniques. The first is a relatively obscure method that has seen little practical application, yet it has significant potential for mapping the paths of possible futures for which there are high levels of uncertainty [20]. The second originated as a way to represent social scientific knowledge through directed graphs, and has since become a common method for scenario analysis in multi-organizational contexts [51]. The third extends the second by making those methods computable for quantitative forecasting, but also has practical uses in a large number of applications across various other domains. Each of these techniques offers insight that contributes to the holistic framework proposed here for forecasting transformative AI.

Scenario network mapping (SNM) is a qualitative scenario technique that was proposed to improve upon existing methods by including a substantially larger number of scenarios, each of which forms a portion of a particular pathway of possible events [52]. This results in a network-like structure which is easily updated in the future with the addition, removal and repositioning of scenarios and their interactions in light of new information. A key feature to SNMs is their reliance on the holonic principle which implies that a scenario can also be decomposed into more scenarios. Following the development of the scenario map, the scenarios can be refined further using causal layered analysis techniques [53]. This technique benefits from larger groups of experts, because the structure of the network becomes more comprehensive with iterative refinement. In a typical SNM scenario building workshop, several hundred possible scenarios are generated, which are then typically reduced to 30–50 plausible scenarios that are used to create the scenario map. Due to this ability to accommodate a large number of plausible scenarios, we see potential for this method for its intended purpose as well as the potential for some derivative of it to be effectively used to identify a large number of possible paths to HLAI (here, we denote the later SNM with an asterisk).

Axelrod first introduced cognitive maps in the 1970s to represent social scientific knowledge with directed graphs [54]. His work has since been extended to a variety of applications including scenario analysis. However, the psychological notion of cognitive maps—people’s representations of their environments and mental modes—comes from two decades earlier [55]. Cognitive maps are effective for facilitating information structuring, elaboration, sequencing and interaction among participants or stakeholders [51]. They are sometimes thought of as causal maps because of the causal network of relationships represented in the nodes and edges. Here nodes can be thought of as scenarios, and the edges describe the causal relationships between them.

Fuzzy cognitive map (FCM) modeling is another hybrid scenario technique that can better integrate expert, stakeholder and historical data through the development of scenarios that assist in linking quantitative models with qualitative storylines [50,56]. FCMs were first proposed by Kosko in the 1980s as a means for making qualitative cognitive maps—used for representing social scientific knowledge [54]—computable by incorporating fuzzy logic. While effective for scenario analysis, FCMs are used generally for decision making and modeling complex systems, and they have a wide variety of applications in multiple domains ranging from online privacy management to robotics [57]. Simply, we can think about FCMs as weighted directed graphs wherein the nodes are fuzzy (i.e.,

they take a continuous value from zero to one rather than a discrete value) and representative of verbally described concepts while the edges are representative of causal effects.

2.3.2. Using Expert Opinion for Scenario Analysis

Virtually all scenario analysis techniques use expert opinion in some way, and there are various ways in which expert opinion is elicited for scenario generation. These techniques include interviews, panels, workshops and the Delphi technique [20]. Many times specific techniques rely directly on the methods for elicitation of expert opinion being employed. For example, the proprietary Interactive Cross-Impact Simulation (INTERAX) methodology relies the generation of a large database of the use of an ongoing Delphi study with close to 500 experts to maintain and update a database of approximately 100 possible events and roughly 50 trend forecasts. Based on six case studies, List suggests that for creating SNMs four half-day workshops with 20 experts is roughly optimal [58]. However, some techniques do not rely specifically on one method for elicitation of expert opinion. FCMs can be developed using expert panels, workshops or interviews. In the case of using interviews, where combining expert opinions is required, all experts' opinions can be treated equally or expert opinions can be weighted based on some assessment of confidence in expert's judgement [56].

2.4. AI Forecasting

The study of forecasting AI and HLAI is in its nascency, and much of the work has relied on expert surveys. The oldest of these dates to a survey conducted in 1972 at a lecture series at the University College of London [59]. Since 2006 12 more surveys have been administered [60]. Such surveys have been used to generate forecasts in the form of timelines. The most recent work has aggregated probability distributions collected from participants [10,61,62]. While the collection and aggregation of probability distributions from experts is an improvement upon previous studies on the topic, there remain many shortcomings in trying to quantify long-term forecasts from surveys of expert opinion, the foremost perhaps being the questionable reliability of experts [25].

The most rigorous of expert survey studies include four particular surveys which have been conducted since 2009, all pertaining to notions of artificial general intelligence (AGI). (Here, we consider human-level machine intelligence, high-level machine intelligence, human-level artificial intelligence and other similar ideas as notions of AGI.) The first of these surveys was conducted at the 2nd Conference on Artificial General Intelligence and found that the majority of experts believed that HLAI would be realized around the middle of the 21st century or sooner [63]. The study also found disagreement among experts concerning the risks involved with AGI and the order of certain milestones (different human-level cognitive tasks) leading to the development of AGI. The next of these studies consisted of a survey that was distributed among four groups of experts at the conference on Philosophy and Theory of AI in 2011, at the AGI-12 conference, to members of the Greek Association for Artificial Intelligence and to the top 100 authors in artificial intelligence by number of citations in May 2013 [64]. This survey questioned participants as to when they expected high-level machine intelligence (HLMI) to be developed, and reported the experts to give a 50% chance of HLMI being developed between 2040 and 2050. These experts further indicated that they believed superintelligence would be created between 2 and 30 years after the emergence of HLMI. Slightly over half of them believed that this would be a positive development while roughly 30% expected it to have negative consequences.

The next survey solicited the primary authors of the 2015 Neural Information Processing Systems (NeurIPS) conference and the 2015 International Conference on Machine Learning (ICML) [61]. This study questioned participants on their forecasts of HLMI, but also included questions about a large number of specific tasks. All forecasters were asked for 10%, 50% and 90% probabilities, which effectively elicited a probability distribution from each. This was not new, but the analysis, including the aggregation of these probability distributions, was novel in the context of AI forecasting. The results indicated a median of 45 years until the development of HLMI, but, interestingly, a median of 120 years

before all human jobs would be automated. The study also found Asian participants to have much earlier predictions than Europeans and North Americans.

The most recent expert survey was solicited at the 2018 International Conference on Machine Learning, the 2018 International Joint Conference on Artificial Intelligence and the 2018 Joint Conference on Human-Level Artificial Intelligence [10]. Rather than focusing on notions of AGI, this study elicited five forecasts for different levels of transformative AI. It also included calibration questions—the first expert survey in the context of AI forecasting to do so. While the forecasts were closely aligned with the previous study, an improved statistical model was used. The use of a naïve calibration technique improved the explainability of the variability in the statistical model for the most extreme transformative AI forecasts. The results also indicated that forecasts from researchers at the HLAI conference were more precise and that this group exhibited lower levels of uncertainty about their forecasts.

A number of meta-analyses of AI forecasting studies have also been conducted. In 2012 and 2014, Armstrong and Sotala and Armstrong et al. assessed previous timeline predictions that had been incorrect [65,66]. They proposed a decomposition schema for analyzing, judging and improving the previous predictions. Muehlhauser has also conducted examinations of timelines and previous AI forecasts [67,68]. His studies offer the most comprehensive discussion of timelines for notions of AGI prior to the surveys conducted over the past decade. Regarding timelines, Muehlhauser concludes that we have learned very little from previous timelines other than the suggestion that it is likely we achieve AGI sometime in the 21st century. He further explores what we can learn from previous timelines and concludes with a list of ten suggestions for further exploration of the existing literature.

AI Impacts (www.aiimpacts.org) is a non-profit organization that is commonly thought to be the leading AI forecasting organization. It has conducted significant work discussing techniques, curating related content and organizing previous efforts for forecasting HLAI, among other research and curation efforts that are aimed at understanding the potential impacts and nature of HLAI. AI Impacts has contributed significantly to practical forecasting knowledge, even leading a major AI forecasting study in 2016 [61].

Recent work by Amodei and Hernandez presented a trendline for the increase in training costs for major milestones in AI progress between 2012 and 2018 [69]. This trendline depicted exponential growth for the increase in the amount of training time required for achieving selected AI milestones; the training time doubled every 3.5 months. However, several critiques of this have emerged [70,71], the most compelling being that from a purely economic perspective the trend was unsustainable for a long period; the exponential rate for training costs was significantly greater than the exponential decrease in costs of compute. Despite these fundamental challenges to the trend, AI experts generally expect the trend to continue for at least 10 years [10]. While not receiving as much visibility, other efforts have been made to plot or collect relevant data to measure the progress of AI research [72–75]. Despite these efforts, the best technology indicator given the criteria previously discussed may be that of Amodei and Hernandez.

While most practical work on AI forecasting to date has relied on expert surveys and extrapolation, there are several important exceptions. Zhang and Dafoe recently conducted a large-scale survey of non-expert opinion that was intended to assess the opinions of the American public regarding AI progress [62]. Another study that was conducted in 2009 used the technology roadmapping technique in an attempt to create a roadmap for the development of HLAI [76]. The results of this workshop depicted expected milestones on the path to HLAI arranged in a two-dimensional grid of individual capability and sociocultural engagement. While organizers of the workshop were disappointed in what they perceived as a failure of the workshop to generate a straightforward roadmap [77,78], arguably 50% or greater of the tasks have been completed [79]. More recently, the Association for the Advancement of Artificial Intelligence and the Computing Community Consortium have completed a 20-year Roadmap for AI research [80]. This roadmap was less ambitious than the earlier attempt, and focused on three major themes: integrated intelligence, meaningful interaction and self-aware learning. The AI Roadmap Institute (www.roadmapinstitute.org) has also been created to study,

create and compare roadmaps to AGI. Although the institute's efforts have resulted in the development of a roadmap, however, it does not concern technical elements as much as social elements.

Another relevant body of research concerns risk analysis. Significant work has been conducted regarding existential risks, and, particularly, the risks posed by AI and superintelligence. (For the purposes of forecasting we do not consider superintelligence, an intelligence explosion or their ramifications [81]). In 2017 Baum created a comprehensive survey of AGI projects that includes a mapping of all relevant stakeholders at the time [82]. Barrett and Baum conducted two studies during 2017, one which focused on the use of fault trees and influence diagrams for risks analysis, and the second which considered expert elicitation methods and aggregation techniques as well as event trees, including a probabilistic framework [83,84]. Additionally, in 2017, Baum et al. examined techniques for modeling and interpreting expert disagreement about superintelligence [85].

Recent work evaluated the methods currently used to quantify existential risk, considering both statistical and judgmental forecasting methods [27]. This study found that while there were no clear 'winners,' the adaptation of large-scale models, the Delphi technique and individual subjective opinions (when elicited through a methodologically rigorous process) had the highest potential. Furthermore, the authors concluded that surveys met all of the criteria to an acceptable degree and that fault trees, Bayesian networks and aggregated expert opinion were all well suited for quantifying AI existential risks. Prediction markets and superforecasting were not found to be especially suitable in general, or for AI risks specifically.

Other recent work by Avin has considered AI forecasting from the futures studies perspective including consideration of the use of scenario planning and wargaming as well as standard judgmental and statistical methods [86]. Wargames (a.k.a. professional role-playing games or government simulation games) are particularly promising as they can be used for informing difficult and complex strategic policy decisions [87]. As mentioned earlier, wargaming can serve two valuable purposes in preparing organizations for futures involving a large degree of uncertainty—training and research—and has been suggested by Avin (in the form of an AI scenario role-play game) as a valuable tool for both of these purposes in the AI strategy field. Furthermore, wargaming can be used in a model-game-model analysis framework to iteratively refine different models for how certain future scenarios may unfold [88]. The work of Beard et al., Barrett and Baum, and Avin represent the only known work in the literature to explore the possibilities for judgmental forecasting techniques (other than surveys) and scenario planning techniques for AI strategy purposes in any depth [27,83]. (We note that there are ongoing efforts to use and improve prediction markets for AI forecasting as well as to develop a new type of forecasting platform for AI forecasting.)

Assessing progress in AI is crucial in order to use extrapolation or other statistical forecasting techniques that require historical data. Consequently, a substantial amount of work has considered theoretical aspects of assessing and modeling AI progress for different ends [89–91]. This discussion focuses on some recent efforts and other notable contributions. In 2018, Martinez-Plumed et al. proposed a framework for assessing AI advances using a Pareto surface which attempted to account for neglected dimensions of AI progress [92]. More recently, in 2019, Martinez-Plumed and Hernandez-Orallo built on previous work on item response theory to propose four indicators for evaluating results from AI benchmarks: two for the milestone or benchmark; difficulty and discrimination, and two for the AI agent; ability and generality [93]. Hernandez-Orallo has written extensively about measures of intelligence intended to be useful for intelligences of all substrates that allow for our existing anthropocentric psychometric tests to be replaced with a universal framework [94]. In contrast to these studies, a measure of intelligence has been proposed by Riedl that is based on creativity [95]. The topic has also garnered mainstream attention with a workshop being dedicated to it in 2015 [96], and work on it being featured in *Nature* in 2016 [97]. Although there is no consensus on how to measure AI progress or intelligence, it is clear that simple measures which can be represented in a small number of dimensions are elusive.

Work by Brundage has attempted to develop a more rigorous framework for modeling progress in AI [6]. In it he suggests that this type of rigorous modeling process as being a necessary precursor to the development of plausible future scenarios for aiding in strategic decision making. To these ends he proposes an AI progress modeling framework that considers the rate of progress in hardware, software, human input elements and specialized AI system performance. In this work, Brundage considered indicators and statistical forecasts as being fundamental for modeling AI progress. However, later efforts by Brundage began to try and integrate scenario planning and judgmental forecasting techniques into formal models (i.e., agent-based models and game theory) [98]. While this later work did not result in the proposal of a rigorous framework like his earlier effort, we believe that it indicates that the integration of various techniques is necessary for adequately modeling and forecasting AI progress. Moreover, it was successful in identifying numerous challenges posed by such an integration. The proposed framework here draws from this previous work in attempts to address these challenges.

In the AI governance research agenda, Dafoe discusses the notion of mapping technical possibilities, or the technical landscape, as a research cluster for understanding possible transformative futures [7]. He also notes the important role of assessing AI progress and modeling AI progress. Separately, he discusses AI forecasting and its challenges and also includes a desiderata for forecasting targets. This paper addresses the issues of generating a mapping, and the task of forecasting events comprising this mapping to the greatest degree that we are able to, as the AI governance research agenda prescribes.

Table 1 compares existing studies to illustrate to readers the focus on surveys and the lack of focus on alternative techniques (it also demonstrates the little amount of work existing in the literature). The earlier work of Brundage seen in Table 1 and discussed in the previous paragraph is the only other known work to consider an integrated and rigorous methodical approach to the specific problem of AI forecasting. However, Brundage did not consider any of these techniques from a forecasting perspective in the way we do. Particularly, his work focuses on applying these techniques directly to a model of AI governance. This study goes further by building on the need for a mapping described by Dafoe and by considering a broader range of forecasting and scenario analysis techniques than previous work to develop a holistic forecasting framework.

Table 1. A comparison of surveys and alternative studies previously conducted on AI forecasting.

Study	Year	Type	Results	Conclusion (Median yrs)
AI Forecasting Surveys				
Baum et al.	2011	Expert (HLAI)	Statistical	Experts expect HLAI in coming decades, much disagreement
Grace et al.	2016	Expert	Probabilistic	45 yrs 50% chance HLAI, Significant cognitive dissonance
Gruetzemacher et al.	2019	Expert (HLAI/AI)	Probabilistic	50 yrs 50% chance HLAI, Type of expertise is significant
Müller and Bostrom	2014	Expert	Statistical	2040–50 50% chance HLAI; <30 yrs to superintelligence
Zhang & Dafoe	2019	Non-expert (Americans)	Probabilistic	54% chance of HLAI by 2028, support AI, weak support HLAI
Other AI Forecasting Studies				
Amodei and Hernandez	2018	Extrapolation	Trendline	Compute required for AI milestones doubling every 18 months
Armstrong and Sotala	2012	Comparative Analysis	Decomposition schema analysis	Expert predictions contradictory and no better than non-experts
Armstrong et al.	2014	Comparative Analysis	Decomposition schema analysis	Models superior to judgment, expert judgment poor, timelines unreliable
Brundage	2016	Methods	Modeling Framework	A framework for modeling AI progress

Table 1. Cont.

Study	Year	Type	Results	Conclusion (Median yrs)
Muehlhauser	2015	Comparative Analysis	Generalization	We know very little about timelines, accuracy is difficult
Muehlhauser	2016	Historical Survey	Suggestions	Future work ideas, AI characterized by periods of hype/pessimism

2.5. Summary of the Related Literature

There are generally thought to be two types of forecasting techniques: judgmental and statistical. Statistical methods are typically preferred when data is available; however, in cases for which data does not exist, is missing or for which there are other inherent irreducible uncertainties, judgmental techniques are commonly the best or only options. AI forecasting falls into the later of these categories. Work from Brundage has previously proposed a general framework for modeling AI progress [6], and later work attempted to integrate scenario analysis, expert judgment and formal modeling [98]. Although most previous studies using judgmental techniques have used expert surveys, there is new evidence that other techniques are more appropriate for this problem [27]. Other potentially valuable techniques, such as tech mining, bibliometric analysis or mapping the technical possibilities have been suggested but have not been attempted in the literature. This study goes further than previous work by considering a holistic framework which attempts to use statistical techniques as best as possible, and to augment their use by including judgmental techniques and scenario analysis techniques. We ultimately take a step beyond forecasting to suggest exercises for strategic management and planning.

3. Judgmental Distillation Mapping

Section 2.3.1 highlighted three scenario analysis techniques that have mapping qualities. We refer to these techniques collectively as scenario mapping techniques due to two significant properties they share: (1) they do not have a strict limit on the number of scenarios they can accommodate and (2) they represent the scenarios as networks with directed graphs (i.e. maps). Although only three have been identified, other approaches are possible. Here, we draw from the existing techniques to propose a new scenario mapping technique which also exhibits the same mapping characteristics as the techniques described in Section 2.3.1. We refer to the proposed technique as judgmental distillation mapping (JDM). Figure 1 depicts a diagram of the JDM process.

Figure 1. The judgmental distillation mapping technique. The technique is flexible and can be thought of as generally being comprised of iterative rounds of questionnaires and interviews intended to isolate a scenario map for which forecasts are generated through Monte Carlo simulation.

The map created (see Figure 2) is distilled from larger input maps (we use map here generically to refer to the combination of tech mining and historical indicators) comprised of both historical or tech mining data, and scenarios developed either by previous rounds of the judgmental distillation process, through interviews or through a technical scenario network mapping workshop (i.e., a scenario network mapping solely for mapping paths to notions of AGI [99]). The scenario map (i.e., the graph) is equivalent in characteristics to that of an FCM, with advanced technologies being represented as nodes. The input nodes represent technologies for which forecasters believe it tractable to use existing forecasting techniques to forecast. The 2nd order and greater nodes in the maps cannot be forecast directly using powerful, traditional techniques such as Delphi, prediction markets or superforecasting. However, these methods are suitable for the first order nodes as long as they are used in a fashion that generates the probability distributions that are necessary for computing the timelines for higher order technologies. These timelines are generated using Monte Carlo simulation and the causal relations between technologies, as determined by expert judgment given the input data. As the final outcomes of transformative AI technology are unknown (possible outcomes include HLAI, comprehensive AI services or AI generating algorithms [100]), the resulting map is able to accommodate a variety of outcomes. Figure 2 depicts a possible result of the JDM process. (This figure is not intended as a forecast, but rather as an example of what JDM could result in. Input distributions are randomly assigned using a normal distribution as opposed to aggregated, adjusted results from JDM.)

Figure 2. Cont.

Figure 2. (a) A hypothetical judgmental distillation map is depicted. White ovals are inputs and light grey ovals are next generation (2nd order) technologies. General intelligence is depicted in a stacked fashion to indicate the possibility of future technological scenarios in the model to be realized through the combination of different paths (i.e., adheres to the holonic principle). The links in the figure are representative of causal relationships and the weights for these links correspond to the strength of these relationships. Note that this figure is not intended to be a forecast, but rather an example of what the JDM process could result in. Input distributions are randomly assigned using a normal distribution. Actual input distributions would not be based on a normal distribution and would be aggregated from expert opinion rather than parameterized distributions. (b) A histogram depicting the results of a Monte Carlo simulation for the next generation adaptive learning technology. (c) A histogram depicting the results of a Monte Carlo simulation for the next generation natural language understanding (NLU) technology. (Monte Carlo simulation is used to generate the distributions found in b and c. A notebook for computing these distributions can be found here: www.github.com/rossgritz/research/.)

JDM is a resource intensive technique that requires a substantial degree of expertise from the forecaster(s) as well as a large number of participating experts. The primary burdens of decomposition and aggregation fall to the facilitator of the process, and, as noted, this provides significant opportunity for facilitators to exercise their own judgment. If not making all input data available to all experts, substantial effort would be required in delegating portions of the input data, and subsequently developing individualized interview questions and questionnaires for the participating experts. To obtain the best results, it is likely best to employ a team for the entire JDM process. Moreover, the best results from the process may be achieved with long periods dedicated to judgmental distillation.

There are three primary inputs for the JDM process: historical data and statistical forecasts, judgmental data and forecasts, and scenarios. Historical data, statistical forecasts and scenarios should initially be in the form of mappings, but such inputs will frequently be deconstructed by the forecasters into forms easily digestible for analysis. Efforts to avoid information overload should be prioritized by forecasters when presenting the information to experts. The questionnaire and interview questions involve asking experts to respond to or comment on scenarios, statistical forecasts, judgmental forecasts and the relationships between these input items. While an example of data that could be shown to interview candidates is shown in Figure 3 and described below, the majority of the JDM process may still be comprised of questions building on qualitative input data rather than quantitative input data.

Figure 3. This depicts a simple extrapolative forecast of a social indicator. This is an example of the type of quantitative information that can be provided to experts for adjustment, distillation and aggregation. When presented with this, experts could be asked whether they agree or disagree that this extrapolation is reasonable. If they disagree, they would be asked to explain how they disagreed and what they thought was a reasonable trend for the indicator presented in the figure. Based on these responses, they may also be questioned about economic realities governing the behavior of this indicator and whether they believed it was possible, even over a substantially longer timeframe, for these economic factors to be altered such that this the indicator may ultimately hit some of the major milestones depicted. They may also be asked questions raising concerns identified by other experts or questions as to why or why not AI research should be analogous to nuclear physics or rocket science. Careful consideration about the indicators and the questions to ask would be determined by the forecaster, or by a forecasting team.

Table 2 is included below to enable an easy comparison of the scenario mapping techniques for readers. It is inserted here so that the newly proposed method of JDM can be included. Therefore, it demonstrates how the new method just described is the only scenario mapping method that is capable of producing probabilistic forecasts for a large number of complex scenarios. The table also indicates that this increased value does come at the cost of substantial resources and a large number of experts (this is discussed further in the discussion section). These factors increase the practical applicability of the method significantly. It can also be noted here that SNM is another useful method for complex cases with large numbers of possible scenarios. Specifically, SNM is useful for mapping the paths to AGI qualitatively while JDM is better suited for generating probabilistic forecasts for AGI and other transformative AI technologies. Overall, (considering the AGI-SNM workshopping technique under development) this table clearly demonstrated the significant practical advantages of the newly proposed methods over the other methods grouped into the scenario mapping class of techniques. (While these new techniques are clearly better for the purposes of forecasting AGI and transformative AI, they may also be useful for other forecasting applications, e.g., for forecasting issues related to existential risks, for forecasting issues related to other complex technology development or for forecasting the broader progress of different domains of scientific study. We do not discuss these options here, but we do suggest that future work consider this broader range of alternate applications.)

Table 2. A comparison of scenario mapping techniques.

Scenario Mapping Techniques					
Technique	No. Scenarios	Quantitative	Qualitative	Strengths	Weaknesses
Scenario Network Mapping (SNM)	30 to 50	No	Yes	Complex cases with large numbers of scenarios	Time consuming and requires 15–20 experts
Cognitive Maps	8 to 24	No	Yes	Useful in multiple organization contexts, rapid workshop development	Weak scenario development, lack of rigor in method
Fuzzy Cognitive Map (FCM)	8 to 18	Yes	Yes	Flexible method, quantitative and qualitative elements, aggregates cognitive maps	Limited quantitative value, limited judgmental value
Judgmental Distillation Mapping (JDM)	6 to 30	Yes	Yes	Complex cases that require probabilistic forecasts	Resource intensive and requires diversity of experts

4. A Holistic Framework for Forecasting AI

JDM was developed to integrate a large variety of forecasts into a single mapping that includes probabilistic timelines for its components. However, these results are more model than narrative and they focus on technological developments rather than economic, political, social or resource-related factors. Moreover, JDM is a technique rather than a broader solution for forecasting, planning and decision making on a continuing basis. The proposed holistic framework uses JDM to leverage flexible combinations of powerful forecasting techniques in an attempt to provide a comprehensive solution.

The framework is depicted in Figure 4. In this figure, inputs are depicted as rectangles, required inputs are depicted as ovals and actionable forecasts are depicted as circles. As depicted in the legend, the elements of the framework can be thought to comprise three distinct groups of processes: input forecasts, JDM and strategic planning. The inputs are comprised of traditional forecasting and scenario analysis/mapping techniques. JDM is modified for the framework to generate two outputs, one directed back at the next iteration of inputs and the other directed at the strategic planning processes. These strategic planning processes then build on JDM forecasts by considering economic, political and technological aspects with both traditional scenario analysis techniques as well as a powerful existing scenario mapping technique. A strategic AI scenario role-playing game can be used for training as well as scenario refinement.

The inputs to the framework are illustrated in Figure 4, and are consistent with the input requirements described for JDM. The quality of the inputs is expected to be strongly correlated with the number of experts, the time requirements and the input forecast quality of JDM. Therefore, we anticipate tech mining, indicators (tech and social), interviews and survey results to all be relatively essential for obtaining a reasonable output given a reasonable amount of resources. The non-essential inputs depicted are SNM, the Delphi technique, superforecasting and prediction markets. The SNM input (scenario network mapping*) is not equivalent to the actionable SNM, but is a workshop technique focused on mapping the paths to strong AI [99] that uses a highly modified form of SNM specifically developed for this purpose. Alternately, the form of SNM used for strategic planning is consistent with the original intentions of the technique and is discussed at length below with the discussion of the strategic planning elements of the framework.

Figure 4. The proposed holistic framework for AI forecasting. Rectangular boxes denote inputs, ovals denote required inputs and circles denote actionable forecasts. Inputs to the framework must include scenarios and a mapping of indicators, however, the specific choice of these and the methods for obtaining them are flexible.

JDM was described thoroughly in the previous subsection. However, in this discussion we considered only the output of the mapping and its timelines. When incorporated into the framework, JDM assumes the dual role of both generating forecasts and informing future component forecasts for iterative application of JDM in the holistic framework. In this way, the framework represents an ongoing forecasting process in which the judgmental distillation process has two objectives and two outputs; one for planning and one for continuing the forecasting process. The figure depicts the first output as moving left for informing actionable strategic planning techniques, and the second element as directing qualitative information right to be merged with updated indicators for developing new targets, forecasts and the next iteration of JDM. This second role performed by JDM, of providing feedback for future iterations, also has the effect of refining the previous forecasts and forecast targets. Since the process of completing an iteration of JDM in the proposed framework is resource intensive, it may be realistic to iterate over longer time periods, e.g., on an annual basis. This may be more amenable to expert participation because the frequency would be less burdensome to the experts, and with a relatively modest incentive (e.g., a gift card or lodging reimbursement for adding a workshop to an annual conference itinerary) participation may pose less of a challenge than other elements of the framework.

The input forecasts and the JDM process can be thought to work in tandem to produce an updated mapping with timelines on a continuing basis. It is likely that this cyclic pair of processes is sufficient to satisfy the requirements of the AI strategy community for a mapping and timelines [7]. However, we do not have to stop here. The purpose of forecasting is to inform decision makers such that they can make the best decisions, and, in order to do this we can draw from the methods discussed in the literature review to extend the JDM results so that they are most effectively used. The weakness of the mapping and timelines resulting from JDM is their heavy focus on the technology. To make the best strategic decisions, numerous factors must be considered (e.g., economic, political and technological). As discussed earlier, scenario analysis techniques are used to incorporate such factors in the planning and decision-making processes. Despite being strongly influenced by scenario analysis techniques, JDM does not include consideration of political factors, social factors or resource-related factors. Therefore, the strategic planning portion of the framework builds on the mapping and timelines for future AI technologies produced from JDM by considering economic, political, social and resource-related factors. It does this by two methods; one for high-level planning (intuitive logics scenarios) and another for exploring granular scenarios (scenario network mapping scenarios). No AI experts are required for these strategic planning elements in the framework. While we only consider the use of scenario planning to improve upon the results of the technology maps and timelines generated from JDM, it is also possible to extend the framework to incorporate forecasts of economic and political events into a separate JDM process keeping the technology map fixed. This possible dual use of JDM underscores the power of the holistic forecasting framework.

The use of intuitive logics scenario planning is inspired by their extensive record of success in business applications. Due to their widespread use and popularity, such scenarios may be more acceptable for use by traditional policy professionals not familiar with AI strategy or advanced scenario planning and technology forecasting methodology (their reliance on narratives makes them more palatable for such persons). In this case, the intuitive logics scenario planning technique would be used as intended to address and plan for uncertainties that are not implicit in the forecasts. This will likely lead to three or four high-level scenarios which can be used for guiding planning and decision-making processes directly. Scenario network maps may have some advantages over intuitive logics scenarios, however, they each can play important roles. Intuitive logics scenarios may be suitable for public dissemination whereas scenario network maps may be too granular and include sensitive information that may not be suitable for public release. Because intuitive logics scenarios may be more appropriate for politicians or other stakeholders who are not familiar with more advanced scenario planning techniques, they are sufficient for official reports from institutions using any holistic forecasting framework. When such scenarios are substantiated by a rigorous forecasting methodology, as discussed here, they can be much more effective for affecting public policy decisions.

SNM is also used here, i.e., the strategic planning context, in the manner that it was intended. However, the workshop technique will need to be modified from that detailed in the SNM manual [101] in order to incorporate the technology map generated from JDM. One of the unique advantages of SNM in this application is its use of the holonic principle. The holonic principle enables the deconstruction of complex scenarios even further. Therefore, Figure 4 includes a self-referential process arrow to indicate the possibility of continuing the scenario decomposition process to the desired level. This could be very useful due to the complex nature of this unique wicked problem [102].

JDM is also a flexible method, and within the holistic framework it can be used to forecast a large variety of AI forecasting targets. The output map would typically be expected to be similar to an FCM, comprised of somewhere between six and twenty nodes. However, there is no fundamental limit on the number of nodes in the map and it could be possible to retain the holonic principle from an input SNM through the distillation process to the output (or to use the holonic principle during the distillation process). Therefore, JDM could be used to forecast automatability of classes of jobs and even particular jobs or tasks with in specific jobs by means of judgmental decomposition. The details of such a process are not discussed at length here, but it is important to realize that as a forecasting

project proceeds to further decompose its targets the required resources would continue to increase. Therefore, while it may be possible to use the framework for forecasting the automation of individual tasks, it is likely not a reasonable pursuit for most organizations due to resource constraints.

5. Discussion

5.1. Strengths and Weaknesses

One of the most obvious weaknesses of both JDM and the proposed framework are the heavy reliance of each on the use and elicitation of expert opinion. These methods may prove difficult to apply when access to experts is limited or biased (as in a single organization). Moreover, the resource requirements may be quite costly in the need for forecasting expertise. The JDM process requires a substantial amount of analysis on the part of the forecaster(s) for deconstructing, creating individualized expert questions and aggregating expert opinion into a single scenario map of AI technologies. The process, as envisioned here, is most likely better suited for teams when working on projects of any reasonable scale. However, the method proposed is flexible and could be revised so as to maintain the holistic framework while reducing reliance on expert judgment. It is unclear whether this would be desirable or not, but it is worth further examination.

The reliance of the proposed method and framework on expert judgment is also one of their greatest strengths. The literature review indicated that for forecasting problems concerning large degrees of uncertainty or for forecasts of rare and unprecedented events, statistical forecasting techniques do not suffice, and judgmental forecasting techniques are required [14,21]. Furthermore, it has been suggested that only those working closely on advanced AI technologies such as AGI may be qualified to make forecasts for such technologies [103]. However, the framework and method proposed here do not go so far as to remove all elements of statistical or data-based forecasting. Rather, we believe that all resources should be used as best as possible. Therefore, the holistic perspective focuses on judgmental techniques while using data-based statistical forecasting techniques to inform them. The framework and method are both inspired by a mixed methods approach to forecasting that uses both qualitative and quantitative judgmental methods. This mixed methods approach to the technique and the framework is another one of its strengths.

5.2. Implications for Practice

Implications for practice are straightforward and have been discussed to some degree in previous sections. However, they raise important questions about the feasibility of practical applications of this technique and framework. For one, the technique is resource intensive and requires a large number of experts for virtually all of the process. Skeptics may see the issue of expert involvement to immediately render the method inviable, and while we believe such a perspective may be extreme, the number or required experts is a credible challenge that should be addressed. There are several ways to consider soliciting experts for participation and they depend on the nature of the organization pursuing the forecast. Academic organizations may have more trouble incentivizing experts while organizations like the Partnership for AI, or the companies which comprise it, may be able to leverage member organizations' or employees' cooperation to obtain expert opinion. It is also likely that motivated and well-funded non-profit organizations (e.g., The Open Philanthropy Project) could effectively solicit expert opinion by means of paying experts appropriately for their time. Another considerable challenge is obtaining an appropriate sample of those actively working on relevant projects, and, based on the nature of the work being done, it may be desirable to intentionally collect biased samples [103]. (This would be equivalent to weighting work being conducted at certain organizations.)

Perhaps equally as costly for practical implementations of JDM and the proposed framework would be the requirement of forecasting expertise. It may be difficult to maintain full-time forecasting experts on the payroll of any organization, even one created specifically for the task of AI forecasting. Limited mappings could be developed by a single forecasting expert, and this may be sufficient for

demonstrating the viability of the concept. However, the most comprehensive and accurate results would likely be realized with forecasting teams. In such teams it may be more appropriate to primarily retain forecasting experts as advisors for management and consultation while employing early career forecasters for the majority of tasks.

5.3. Implications for Research

AI forecasting is a nascent discipline, and, to date, no unifying document exists. While this work does not intend to be such a document, we do wish to draw attention to the need of the research community for one. This document presents a framework that relies on numerous techniques working harmoniously toward a single goal. Therefore, research to improve the method and framework may be most effective through analysis of some of the contributing elements. Moreover, work is also necessary which explores this framework, improvements or variations thereof, and alternate ways to incorporate judgmental forecasting methods with statistical forecasting methods and scenario analysis methods (e.g., Brundage [98]). Here we outline some suggestions of this type for future work that could be elaborated upon in the form of an AI forecasting research agenda. A structured research agenda with a coherent vision for the forecasting space could act as the sort of unifying document needed for the AI forecasting research space. (The method and framework presented here may contribute to a clear vision for the AI forecasting space, but further input is needed from others with experience in the field to iron out a unified vision.)

This study has illuminated a large number of topics that do not seem to have received appropriate attention thus far in the study of AI forecasting. Other recent work has identified some of these topics as salient [27,86], however, the previous work has not gone so far as to suggest action to motivate progress in future research. To our knowledge, no literature review exists that is equal in scope to the one presented here with respect to AI forecasting (the depth of this literature review leaves much to be desired). We believe that going forward a major priority in the study of AI forecasting is the necessity of a large number of comprehensive literature reviews for narrow topics (e.g. the many techniques discussed here) in the context of how they may be used for the tasks involved in AI forecasting. Saura et al. demonstrate an excellent example of an effective literature review for a related topic that offers a good model for such work [104]. We also see the need for a broad, comprehensive literature review—the literature review here may be a good start, but we argue that a dedicated document is desirable. These suggestions are mentioned first as they may be the lowest hanging fruit but also have significant potential for being very useful.

The literature review here found the body of existing work was lacking in studies that had compared forecasting methods. Of the studies that did, none of these compared superforecasting and none of these considered methods when used for the purpose of technological forecasting or for AI forecasting specifically. Moreover, work considering the Delphi technique found academic work assessing its viability to be lacking due to the excessive use of students rather than professionals and experts. (The Delphi technique is intended specifically for use with experts. Some studies with students have attempted forecasts of things such as college football, for which students may be considered experts, however, the vast majority of these studies did not [21].) Since the literature review was not comprehensive, a focused and more extensive effort may illuminate valuable work that has not yet been uncovered (this illustrates the possible significance that literature reviews can play). Regardless, it is clear that significant future work on methods evaluation and comparison, particularly for the viability of various forecasting techniques in the context AI forecasting, are required in order to best determine how and when the wide variety of methods are suitable in this framework and when they are suitable for AI forecasting purposes more generally. Three methods are depicted in Figure 4 as being optional inputs for JDM in the framework: the Delphi technique, prediction markets and superforecasting. It may be that one of these is indeed superior for the majority of related tasks, or, that they each can serve certain purposes in a balanced capacity to achieve the best results. Priorities for future research include comparing these three methods directly, as well as comparing the suitability of these methods

in the JDM process. Such comparisons are useful both in the context of AI forecasting as well as in other contexts.

Calibration is another topic related to judgmental forecasting techniques that could be helpful for AI forecasting if better understood. Gruetzemacher et al. recently demonstrated an alternative calibration technique (i.e. naïve calibration) that demonstrates the possibility of novel calibration techniques [10]. While calibration is widely used and fundamental to superforecasting techniques, there remains little work on the topic. No work exists to confirm or assess the value that calibration training can have in forecasting, or what level of training is necessary for improving untrained experts' forecasts. Straightforward empirical studies to assess calibration training or different types of new calibration techniques for various judgmental forecasting techniques are likely very valuable. As Gruetzemacher et al. has recently shown, a substantial proportion of AI experts' forecasts are poorly calibrated, such studies to improve and better understand calibration techniques could have a quick and nontrivial return for AI forecasting efforts.

Tech mining and bibliometric mapping have a huge role to play in the proposed framework, and likely in any AI forecasting framework. Brundage mentions the use of bibliometric methods for mapping inputs [6], however, no work is known which has pursued this suggestion. While the foremost priority is likely a review of the related literature, practical work is also desperately needed. It is likely that these techniques must be used to some degree to demonstrate and/or validate the proposed method and framework, but a more extensive examination of these techniques should also be a priority. Examples of such powerful new techniques for language modeling [105], citation analysis [106] and data text mining [107] should be explored for their suitability in this topic (a literature review could likely identify even more). A large body of software also exists for the mapping of science [108], however, each flavor produces a different result. We are uncertain as to what form of results will in fact be useful for judgmental distillation, or for alternate forecasting frameworks, and this is of critical interest for an initial inquiry or for numerous simultaneous inquiries. It may be that several techniques are valid and can be shown to one expert or different experts in the JDM process. It could also be that an interactive mapping platform is most valuable for judgmental distillation in that such a platform could enable active navigation through complex maps in three dimensions. If this is the case, and if the needs for AI mapping are not met with existing software, future work could also be necessary for developing, testing and refining a tech mining based AI mapping platform. Regardless, it seems imperative that work on these topics be prioritized.

Another topic of interest is that of relevant indicators of AI progress. The literature review here discussed a growing body of work in this direction, and this work is certainly desirable moving forward. The framework proposed here takes a slightly different perspective than a substantial portion of the work toward these ends in that it suggests value in a larger number of salient indicators as opposed to a smaller number of indicators. A larger number of indicators is more realistic for the high dimensional space of AI progress, and trying to reduce progress toward broadly capable systems to a small number of vectors ignores the fundamental uncertainty of the technology that we are trying to forecast. The indicator proposed by Amodei and Hernandez met the criteria for technology indicators relatively well [69], as did some of the other metrics that have been developed for measuring AI progress [93]. Ongoing efforts toward the latter are likely sufficient at this time, but efforts to explore indicators like the former, or like the one depicted in Figure 3, while being difficult to identify, should be considered a priority. There are a large range of possible outcomes and any forecasting framework must consider this.

Finally, forecasting targets are another critical topic that should be considered for future research efforts. The desiderata presented by Dafoe is an excellent ideal [7], however, in practice it can be challenging to develop targets. Structured methodologies to develop these targets are highly desirable. Work on such methods is therefore a priority given the significance of including expert forecasts for adjustment in addition to statistical forecasts. For example, workshopping techniques or interview techniques to identify and refine these targets are sorely needed. Targets that suffice may be more

realistic than ideal targets, and a list of minimum criteria may be useful in addition to Dafoe's desiderata. An initial effort to expound on Dafoe's work and to examine its strengths and weaknesses could be a simple yet valuable contribution now. Also of interest are the effects of combining statistical forecasts with judgmental forecasts and aggregated expert opinion forecasts for adjustment, or the effects of combining multiple forecasts of other variations. Work exploring these effects could be examined with or without the use of domain expertise and could have significant implications on how forecasters deconstruct and delegate questions to experts in the judgmental distillation process. Furthermore, the methods of determining forecasting targets' resolution may sometimes be ambiguous and techniques are necessary for objectively resolving forecasting targets.

5.4. Challenges and Future Work

This paper only outlines and describes a new method and a new forecasting framework. Many challenges lie ahead for continuing research on this topic. Foremost, efforts should be undertaken to evaluate and validate both the method and framework proposed here. The method may be possible to evaluate objectively in contexts other than AI forecasting, doing so may be a good step for confirming the viability of the method and the framework, however, evaluation should not be limited to toy cases. An alternate means of validation is to employ the method first in a preliminary fashion, as for demonstrating viability, and then to pursue a full-scale implementation of the method and framework. Results from the former could be used for validating and refining the technique and framework through the inclusion of calibration targets ranging from one to three years. If this was done to pilot the proposed method and framework, validation of the method would be confirmed gradually over three years. If successful early on, then further resources could be justified moving forward if performance persisted. This would also have the added benefits of improving training for AI strategy researchers and professionals and improving the scenario planning capabilities of the strategy community. If timeline forecasts were equal to or less accurate than existing methods, over short time frames, then qualitative assessment of the benefits to the planning process would have to be considered also, and the forecasting framework could be modified. Work is ongoing toward these ends. Other work should also be prioritized that assesses and validates the framework and method in other contexts which may not take as long, or, that works to refine and improve the framework and method. The framework and new method could also possibly be decomposed, evaluated and validated piece-wise to expediate the process. Regardless the path chosen, much difficult work certainly lies ahead.

6. Conclusions

The framework proposed here is not intended to be the solution for AI forecasting. Rather, it is intended to illuminate the possibility of considering a holistic perspective when addressing the unique challenges of AI forecasting. It differs from previous work in its holistic perspective and through the development of a new method for judgmental distillation of the salient features of a diverse group of forecasting techniques. By incorporating expert judgment in addition to historical and mined data, it attempts to address issues of severe uncertainty inherent in AI forecasting, while still harnessing the power of statistical methods and scenario analysis in a novel manner through the proposed framework.

There are several significant novel contributions of this work. First, the paper proposes and outlines a new method for mapping and forecasting transformative AI with judgmental distillation mapping (JDM). Second, the paper proposes and outlines a new framework for forecasting transformative AI that builds on the new method of JDM while incorporating a variety of forecasting techniques in a holistic approach. Finally, the paper approaches the problem of forecasting in a holistic manner by incorporating many competing methods into a single forecasting ecosystem. There are significant social implications because this method and framework, unlike any other approaches, is able to combine complex yet plausible future scenarios with a rigorous methodological foundation. In doing so, it has the potential to compel lawmakers to act on policy recommendations that in the past have

seemed too unrealistic or implausible. Ultimately, the intended beneficiaries of this new approach are lawmakers constituents and all the world's citizens.

Funding: This research received no external funding.

Acknowledgments: We would like to thank Zoe Cremer, David Paradice and Christy Manning for their comments, suggestions and assistance in various aspects of the development of this text.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Roper, A.T.; Cunningham, S.W.; Porter, A.L.; Mason, T.W.; Rossini, F.A.; Banks, J. *Forecasting and Management of Technology*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. Available online: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (accessed on 21 June 2019).
3. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef]
4. Team. OpenAI Five. Available online: <https://openai.com/blog/openai-five/> (accessed on 31 May 2019).
5. Building High-Level Features Using Large Scale Unsupervised Learning. Available online: <https://icml.cc/2012/papers/73.pdf> (accessed on 21 June 2019).
6. Brundage, M. Modeling progress in AI. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
7. Dafoe, A. *AI Governance: A Research Agenda*; Future of Humanity Institute, University of Oxford: Oxford, UK, 2018.
8. Rahwan, I.; Cebrian, M.; Obradovich, N.; Bongard, J.; Bonnefon, J.-F.; Breazeal, C.; Crandall, J.W.; Christakis, N.A.; Couzin, I.D.; Jackson, M.O. Machine Behaviour. *Nature* **2019**, *568*, 477–486. [CrossRef]
9. Duckworth, P.; Graham, L.; Osborne, M.A. Inferring Work Task Automatability from AI Expert Evidence. In Proceedings of the 2nd Conference on Artificial Intelligence for Ethics and Society, Honolulu, HI, USA, 26–28 January 2019.
10. Forecasting Transformative AI: An Expert Survey. Available online: <https://arxiv.org/abs/1901.08579> (accessed on 21 June 2019).
11. Hutter, M. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*; Springer: Berlin/Hiedelberg, Germany, 2005.
12. Minsky, M.L.; Singh, P.; Sloman, A.J. The St. Thomas common sense symposium: Designing architectures for human-level intelligence. *AI Mag.* **2004**, *25*, 113.
13. Drexler, K.E. Reframing Superintelligence. Available online: https://www.fhi.ox.ac.uk/wp-content/uploads/Reframing_Superintelligence_FHI-TR-2019-1.1-1.pdf (accessed on 31 May 2019).
14. Armstrong, J.S. *Principles of Forecasting: A Handbook for Researchers and Practitioners*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Volume 30.
15. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
16. Orrell, D.; McSharry, P. System economics: Overcoming the pitfalls of forecasting models via a multidisciplinary approach. *Int. J. Forecast.* **2009**, *25*, 734–743. [CrossRef]
17. Cooke, R. *Experts in Uncertainty: Opinion and Subjective Probability in Science*; Oxford University Press on Demand: New York, NY, USA, 1991.
18. Aspinall, W. A route to more tractable expert advice. *Nature* **2010**, *463*, 294. [CrossRef]
19. Bradfield, R.; Wright, G.; Burt, G.; Cairns, G.; Van Der Heijden, K. The origins and evolution of scenario techniques in long range business planning. *Futures* **2005**, *37*, 795–812. [CrossRef]
20. Amer, M.; Daim, T.U.; Jetter, A. A review of scenario planning. *Futures* **2013**, *46*, 23–40. [CrossRef]
21. Goodwin, P.; Wright, G. The limits of forecasting methods in anticipating rare events. *Technol. Forecast. Soc. Chang.* **2010**, *77*, 355–368. [CrossRef]
22. Rowe, G.; Wright, G. Expert opinions in forecasting: The role of the Delphi technique. In *Principles of Forecasting*; Springer: Dordrecht, The Netherlands, 2001; pp. 125–144.

23. Arrow, K.J.; Forsythe, R.; Gorham, M.; Hahn, R.; Hanson, R.; Ledyard, J.O.; Levmore, S.; Litan, R.; Milgrom, P.; Nelson, F.D.; et al. The Promise of Prediction Markets. *Science* **2008**, *320*, 877–878. [[CrossRef](#)]
24. Green, K.C.; Armstrong, J.S.; Graefe, A. Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *Foresight* **2007**, *8*, 17–20.
25. Tetlock, P.E.; Gardner, D. *Superforecasting: The Art and Science of Prediction*; Penguin Random House: New York, NY, USA, 2016.
26. Schoemaker, P.J.; Tetlock, P.E. Superforecasting: How to upgrade your company's judgment. *Harv. Bus. Rev.* **2016**, *94*, 72–78.
27. Beard, S.; Rowe, T.; Fox, J. *An Analysis and Evaluation of Methods Currently Used to Quantify Existential Risk*, under review.
28. Sanders, N.R.; Ritzman, L.P. Judgmental adjustment of statistical forecasts. In *Principles of Forecasting*; Springer: Dordrecht, The Netherlands, 2001; pp. 405–416.
29. Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science* **1974**, *185*, 1124–1131. [[CrossRef](#)]
30. Goodwin, P.; Wright, G. Enhancing strategy evaluation in scenario planning: A role for decision analysis. *J. Manag. Stud.* **2001**, *38*, 1–16. [[CrossRef](#)]
31. Wright, G.; Goodwin, P. Decision making and planning under low levels of predictability: Enhancing the scenario method. *Int. J. Forecast.* **2009**, *25*, 813–825. [[CrossRef](#)]
32. Lipinski, A.; Loveridge, D. Institute for the future's study of the UK, 1978–1995. *Futures* **1982**, *14*, 205–239. [[CrossRef](#)]
33. Rea, L.M.; Parker, R.A. *Designing and Conducting Survey Research: A Comprehensive Guide*; John Wiley & Sons: San Francisco, CA, USA, 2014.
34. Watts, R.J.; Porter, A.L. Innovation forecasting. *Technol. Forecast. Soc. Chang.* **1997**, *56*, 25–47. [[CrossRef](#)]
35. Porter, A.L.; Cunningham, S.W. Tech mining. *Compet. Intell. Mag.* **2005**, *8*, 30–36.
36. Phaal, R.; et al. Technology roadmapping—A planning framework for evolution and revolution. *Technol. Forecast. Soc. Chang.* **2004**, *71*, 5–26. [[CrossRef](#)]
37. Duin, P.A. *Qualitative Futures Research for Innovation*; Eburon Academic Publishers: Delft, The Netherlands, 2006.
38. Garcia, M.L.; Bray, O.H. *Fundamentals of Technology Roadmapping*; Sandia National Labs: Albuquerque, NM, USA, 1997.
39. Rip, A. Mapping of science: Possibilities and limitations. In *Handbook of Quantitative Studies of Science and Technology*; Elsevier: Amsterdam, The Netherlands, 1988; pp. 253–273.
40. Tijssen, R.J.; Van Raan, A.F. Mapping changes in science and technology: Bibliometric co-occurrence analysis of the R&D literature. *Eval. Rev.* **1994**, *18*, 98–115.
41. Nagy, B.; Farmer, J.D.; Bui, Q.M.; Trancik, J.E. Statistical basis for predicting technological progress. *PLoS ONE* **2013**, *8*, e52669. [[CrossRef](#)]
42. Mullins, C. *Retrospective Analysis of Technology Forecasting: In-Scope Extension*; The Tauri Group: Alexandria VA, USA, 2012.
43. Brynjolfsson, E.; Mitchell, T. What can machine learning do? Workforce implications. *Science* **2017**, *358*, 1530–1534. [[CrossRef](#)]
44. Van der Heijden, K.; Bradfield, R.; Burt, G.; Cairns, G.; Wright, G. *The Sixth Sense: Accelerating Organizational Learning with Scenarios*; John Wiley & Sons: San Francisco, CA, USA, 2002.
45. Perla, P.P. *The Art of Wargaming: A Guide for Professionals and Hobbyists*; Naval Institute Press: Annapolis, MD, USA, 1990.
46. Roxburgh, C. The use and abuse of scenarios. *Mckinsey Q.* **2009**, *1*, 1–10.
47. Chermack, T.J.; Lynham, S.A.; Ruona, W.E. A review of scenario planning literature. *Futures Res. Q.* **2001**, *17*, 7–32.
48. Gordon, T.J.; Helmer, O. *Report on a Long-Range Forecasting Study*; Rand Corp: Santa Monica, CA, USA, 1964.
49. Schoemaker, P.J. Scenario planning: A tool for strategic thinking. *Sloan Manag. Rev.* **1995**, *36*, 25–50.
50. Van Vliet, M.; Kok, K.; Veldkamp, T. Linking stakeholders and modellers in scenario studies: The use of Fuzzy Cognitive Maps as a communication and learning tool. *Futures* **2010**, *42*, 1–14. [[CrossRef](#)]

51. Soetanto, R.; Dainty, A.R.; Goodier, C.I.; Austin, S.A. Unravelling the complexity of collective mental models: A method for developing and analysing scenarios in multi-organisational contexts. *Futures* **2011**, *43*, 890–907. [CrossRef]
52. List, D. Scenario network mapping. *J. Futures Stud.* **2007**, *11*, 77–96.
53. Inayatullah, S. Causal layered analysis: Poststructuralism as method. *Futures* **1998**, *30*, 815–829. [CrossRef]
54. Axelrod, R. *Structure of Decision: The Cognitive Maps of Political Elites*; Princeton University Press: Princeton, NJ, USA, 2015.
55. Tolman, E.C. Cognitive maps in rats and men. *Psychol. Rev.* **1948**, *55*, 189–208. [CrossRef]
56. Jetter, A.J.; Kok, K. Fuzzy Cognitive Maps for futures studies—A methodological assessment of concepts and methods. *Futures* **2014**, *61*, 45–57. [CrossRef]
57. Papageorgiou, E.I. *Fuzzy Cognitive Maps for Applied Sciences and Engineering: From Fundamentals to Extensions and Learning Algorithms*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 54.
58. List, D. *Scenario Network Mapping: The Development of a Methodology for Social Inquiry*; University of South Australia: Adelaide, Australia, 2005.
59. Michie, D. Machines and the theory of intelligence. *Nature* **1973**, *241*, 507–512. [CrossRef]
60. Grace, K. AI Timeline Surveys. Available online: <https://aiimpacts.org/ai-timeline-surveys/> (accessed on 31 May 2019).
61. Grace, K.; Salvatier, J.; Dafoe, A.; Zhang, B.; Evans, O. When will AI exceed human performance? Evidence from AI experts. *J. Artif. Intell. Res.* **2018**, *62*, 729–754. [CrossRef]
62. Zhang, B.; Dafoe, A. *Artificial Intelligence: American Attitudes and Trends*; University of Oxford: Oxford, UK, 2019.
63. Baum, S.D.; Goertzel, B.; Goertzel, T.G. How long until human-level AI? Results from an expert assessment. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 185–195. [CrossRef]
64. Müller, V.C.; Bostrom, N. Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental Issues of Artificial Intelligence*; Springer International Publishing: Basel, Switzerland, 2016; pp. 555–572.
65. Armstrong, S.; Sotala, K. How we’re predicting AI—or failing to. In *Beyond Artificial Intelligence*; Springer: Pilsen, Czech Republic, 2015; pp. 11–29.
66. Armstrong, S.; Sotala, K.; hEigeartaigh, S.Ó. The errors, insights and lessons of famous AI predictions—and what they mean for the future. *J. Exp. Theor. Artif. Intell.* **2014**, *26*, 317–342. [CrossRef]
67. What Do We Know About AI Timelines? Available online: <https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/ai-timelines> (accessed on 31 May 2019).
68. What Should We Learn from Past AI Forecasts? Available online: <https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/what-should-we-learn-past-ai-forecasts> (accessed on 31 May 2019).
69. AI and Compute. Available online: <https://openai.com/blog/ai-and-compute/> (accessed on 31 May 2019).
70. Interpreting AI Compute Trends. Available online: <https://aiimpacts.org/interpreting-ai-compute-trends/> (accessed on 31 May 2019).
71. Reinterpreting “AI and Compute”. Available online: <https://aiimpacts.org/reinterpreting-ai-and-compute/> (accessed on 31 May 2019).
72. Measuring the Progress of AI Research. Available online: <https://www.eff.org/ai/metrics> (accessed on 31 May 2019).
73. Trends in Algorithmic Progress. Available online: <https://aiimpacts.org/trends-in-algorithmic-progress/> (accessed on 31 May 2019).
74. Constantin, S. Performance Trends in AI. Available online: <https://srconstantin.wordpress.com/2017/01/28/performance-trends-in-ai/> (accessed on 31 May 2019).
75. AI Metrics Data. Available online: https://raw.githubusercontent.com/AI-metrics/master_text/master/archive/AI-metrics-data.txt (accessed on 31 May 2019).
76. Adams, S.; Arel, I.; Bach, J.; Coop, R.; Furlan, R.; Goertzel, B.; Hall, J.S.; Samsonovich, A.; Scheutz, M.; Schlesinger, M. Mapping the landscape of human-level artificial general intelligence. *AI Mag.* **2012**, *33*, 25–42. [CrossRef]
77. Goertzel, B. *The AGI Revolution: An Inside View of the Rise of Artificial General Intelligence*; Humanity+ Press: Los Angeles, CA, USA, 2016.
78. Goertzel, B. *Ten Years to the Singularity If We Really Really Try*; Humanity+ Press: Los Angeles, CA, USA, 2014.

79. Gruetzmacher, R.; Paradice, D. Alternative Techniques for Mapping Paths to HLAI. *arXiv* **2019**, arXiv:1905.00614.
80. Computing Community Consortium (CCC) (Ed.) Townhall: A 20-Year Roadmap for AI Research. In Proceedings of the 33rd Annual Conference for the Association of the Advancement of Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
81. Bostrom, N. *Superintelligence*; Oxford University Press: Oxford, UK, 2014.
82. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070741 (accessed on 31 May 2019).
83. Barrett, A.M.; Baum, S.D. Risk analysis and risk management for the artificial superintelligence research and development process. In *The Technological Singularity*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 127–140.
84. Barrett, A.M.; Baum, S.D. A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. *J. Exp. Theor. Artif. Intell.* **2017**, *29*, 397–414. [CrossRef]
85. Baum, S.; Barrett, A.; Yampolskiy, R.V. Modeling and interpreting expert disagreement about artificial superintelligence. *Informatica* **2017**, *41*, 419–428.
86. Avin, S. Exploring Artificial Intelligence Futures. *J. AI Humanit.* Forthcoming.
87. Parson, E.A. *What Can You Learn from A Game? Wise Choices: Games, Decisions, and Negotiations*; Harvard Business School Press: Boston, MA, USA, 1996.
88. Davis, P.K. *Illustrating a Model-Game-Model Paradigm for Using Human Wargames in Analysis*; RAND National Defense Research Institute: Santa Monica, CA, USA, 2017.
89. Fernández-Macias, E.; Gómez, E.; Hernández-Orallo, J.; Loe, B.S.; Martens, B.; Martínez-Plumed, F.; Tolan, S. A multidisciplinary task-based perspective for evaluating the impact of AI autonomy and generality on the future of work. *arXiv* **2018**, arXiv:1807.02416.
90. Evaluation of General-Purpose Artificial Intelligence: Why, What & How. Available online: http://dmip.webs.upv.es/EGPAI2016/papers/EGPAI_2016_paper_9.pdf (accessed on 31 May 2019).
91. Hernández-Orallo, J. AI Evaluation: Past, Present and Future. *arXiv* **2014**, arXiv:1408.6908.
92. Martínez-Plumed, F.; Avin, S.; Brundage, M.; Dafoe, A.; hÉigeartaigh, S.Ó.; Hernández-Orallo, J. Accounting for the neglected dimensions of ai progress. *arXiv* **2018**, arXiv:1806.00610.
93. Martínez-Plumed, F.; Hernández-Orallo, J. Analysing Results from AI Benchmarks: Key Indicators and How to Obtain Them. *arXiv* **2018**, arXiv:1811.08186.
94. Hernández-Orallo, J. *The Measure of All Minds: Evaluating Natural and Artificial Intelligence*; Cambridge University Press: Cambridge, UK, 2017.
95. Riedl, M.O. The Lovelace 2.0 test of artificial intelligence and creativity. In Proceedings of the 29th AAAI Conference on Artificial Intelligence Workshops, Austin, TX, USA, 25–26 January 2015.
96. Hernández-Orallo, J.; Baroni, M.; Bieger, J.; Chmait, N.; Dowe, D.L.; Hofmann, K.; Martínez-Plumed, F.; Strannegård, C.; Thórisson, K.R. A new AI evaluation cosmos: Ready to play the game? *AI Mag.* **2017**, *38*, 66–69. [CrossRef]
97. Castelvechi, D. Tech giants open virtual worlds to bevy of AI programs. *Nat. News* **2016**, *540*, 323. [CrossRef]
98. Brundage, M. *Responsible Governance for Artificial Intelligence: An Assessment, Theoretical Framework, and Exploration*, 2018, Unpublished.
99. Gruetzmacher, R.; Paradice, D. Mapping the Paths to AGI. In Proceedings of the 12th Annual Conference on Artificial General Intelligence, Shenzhen, China, 6–9 August 2019.
100. Clune, J. AI-GAs: AI-Generating Algorithms, an Alternate Paradigm for Producing General Artificial Intelligence. *arXiv* **2019**, arXiv:1905.10985.
101. List, D. *Scenario Mapping: A User's Manual*; Original Books: Adelaide, Australia, 2006.
102. Gruetzmacher, R. Rethinking AI Strategy and Policy as Entangled Super Wicked Problems. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; p. 122.
103. There's No Fire Alarm for Artificial General Intelligence. Available online: <https://intelligence.org/2017/10/13/fire-alarm/> (accessed on 31 May 2019).
104. Saura, J.R.; Palos-Sánchez, P.; Cerdá Suárez, L.M. Understanding the digital marketing environment with KPIs and web analytics. *Future Internet* **2017**, *9*, 76. [CrossRef]

105. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
106. Cohan, A.; Ammar, W.; van Zuylen, M.; Cady, F. Structural Scaffolds for Citation Intent Classification in Scientific Publications. *arXiv* **2019**, arXiv:1904.01608.
107. Saura, J.R.; Bennett, D.R. A Three-Stage method for Data Text Mining: Using UGC in Business Intelligence Analysis. *Symmetry* **2019**, *11*, 519. [[CrossRef](#)]
108. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. Science mapping software tools: Review, analysis, and cooperative study among tools. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 1382–1402. [[CrossRef](#)]

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).