Journal of Futures Studies
2023,Vol. 28(2) 79-99
DOI: 10.6531/JFS.202312_28(2).0005

Article

How Superintelligence Affects Human Health: A Scenario Analysis

Philipp Koebe[1], Tobias Schillings[2], Jan Oliver Schwarz[3]

[1] *Witten/Herdecke University: Witten, DE, ORCID: 0000-0002-1345-5405*
[2] *University of Oxford, Oxford, UK*
[3] *Technische Hochschule Ingolstadt: Ingolstadt, DE, ORCID: 0000-0002-2995-5308*

Abstract

Population health is a crucial determinant of human prosperity and well-being. Poor health can lead to reduced productivity, poverty, and premature death, with the COVID-19 pandemic underscoring the vulnerability of population health on a global scale. Self-learning algorithms have the potential to improve population health in a sustainable way and bring a paradigm shift to healthcare. We utilize intuitive logic to generate future scenarios in order to address the research question. These scenarios are categorized as either health-promoting or health-damaging, and superintelligence is considered either dominating or non-dominating. We provide strategic implications for each scenario, which can guide policy action in dealing with superintelligence.

Keywords

Superintelligence, Artificial Intelligence, Future Scenario, Public Health, Healthcare

Introduction

Anticipating future developments and assessing potential risks is an essential process for engaging with new technologies and understanding their impact on fundamental components of society (Torres, 2019). Population health plays a crucial role in this process, as it is a necessary condition for the success and continuation of humanity. Significant single events, such as the eruption of the Pompeii volcano (Giacomelli et al., 2003) or the Chernobyl nuclear accident (Cardis & Hatch, 2011), led to significant consequences for the population living in the world at that time, their health, and their actions to prevent future existential threats. In addition to single events, long-term changes that may not immediately manifest can also pose existential threats. Examples include species extinction or climate change, both of which have wide-ranging implications for human health and well-being (Tol, 2020; Whitmee et al., 2015). Monitoring existential risks from an intergenerational justice perspective is a critical lever for securing a livable future (Werther, 2013).

The Importance of Technology for Public Health

The significance of population health cannot be overstated, as it is a fundamental requirement for human prosperity and well-being (Steptoe et al., 2015). The global COVID-19 pandemic has

* *Corresponding author.*
E-mail addresses: philipp.koebe@uni-wh.de (P. Koebe)

Received: 29 September 2022 and Accepted: 6 June 2023

brought attention to the vulnerability of health and the need for measures to prevent future pandemics (Bambra et al., 2020). However, evidence shows that acceptance of risk-reduction measures, such as vaccinations (Lazarus et al., 2021) or restrictive public health behaviors (Kachanoff et al., 2021), varies widely among populations. Technology can play a crucial role in supporting such public health measures, but similarly, societal attitudes vary towards issues around surveillance (Ioannou & Tussyadiah, 2021) and sharing of health data (Huston et al., 2019). Despite these concerns, medical and technological advancements already change how healthcare is delivered around the globe (Breyer & Felder, 2006). Among these advancements, artificial intelligence (AI) appears as the trend that is most likely to fundamentally shape the development of medicine in the 21st century (Tortorella et al., 2020). Therefore, this paper will explore the potential impacts of AI on the area of population health.

Defining Superintelligence

This paper employs the definition of superintelligence according to Bostrom (1998):

> “By a "superintelligence" we mean an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.”

We will consider AI as a specific form of superintelligence that could play a supporting role in future societies (Brundage, 2015). One area where AI algorithms and superintelligences can significantly contribute is healthcare, for example in the development of new diagnostics and therapies to eradicate diseases (Meskó et al., 2018). Thus, superintelligences have the potential to bring great benefits, but they could also cause great harm (Sotala, 2017). They could potentially pose a significant threat to humanity in the future, which requires us to identify and evaluate them now to take preventive measures (Winch & Maytorena-Sanchez, 2011). In this paper, we will use a matrix to rank both known and unknown parameters to assess the impact of superintelligence (Mercer & Trothen, 2021). The chosen uncertainty paradigm is shown in Figure 1.

When we have knowledge of the properties of AI and the effects of its use, we can consider possibility spaces to prevent potential risks. However, since superintelligence or artificial general intelligence does not exist yet, we cannot evaluate its properties or mode of action. However, as the properties and effects of individual AI applications for healthcare services are already known (Ahmed et al., 2020), we can use them to draw conclusions about the potential influence of superintelligence (Batin et al., 2017). While we cannot anticipate the full impacts of its deployment, some of these individual insights can be generalized and used to limit uncertainty. In this paper, we aim to approach these unknown parameters and highlight potential health risks of superintelligence use.

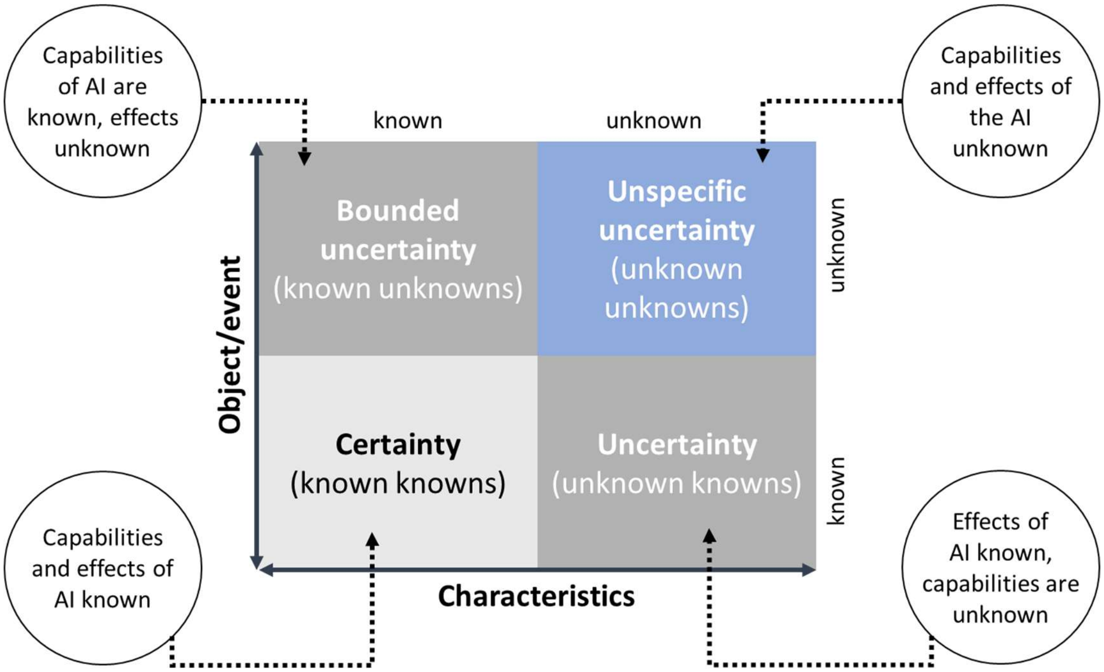

Capabilities of AI are known, effects unknown
Capabilities and effects of the AI unknown
known
unknown
Bounded uncertainty (known unknowns)
Unspecific uncertainty (unknown unknowns)
unknown
Object/event
Certainty (known knowns)
Uncertainty (unknown knowns)
known
Capabilities and effects of AI known
Characteristics
Effects of AI known, capabilities are unknown

Fig 1: Uncertainty paradigms related to AI in the future

Methodology

Scenario techniques are a suitable method for classifying and evaluating future developments, possible shocks and unexpected events (Huss & Honton, 1987; Schoemaker, 1995). Especially under extreme uncertainty, scenarios can provide a way to anticipate the future and foster strategic impetus for action (Schoemaker, 2004; Wright & Goodwin, 2009). This approach enables the development of action guidelines that can help avert existential risks for companies, states, or humanity as a whole. In this work, we use the intuitive logic method to develop future scenarios (Wright et al., 2013). This methodology was first described by Schoemaker and van der Heijden (1992) in the context of strategic planning for the Royal Dutch/Shell company. It identifies two dimensions that are particularly relevant to the future problem and describes four possible scenarios. We follow the methodological approach of Bradfield (2008), as shown in Figure 2. First, we define the problem domain. As a basis for the initial evaluation, the scenario analysis of Reinhart and Greiner (2019) is used which considers the dominance of superintelligences and their positive or negative disposition towards humans. Given the particular challenges and relevance of the topic, we defined global population health as an additional parameter, as explained earlier.

After having identified and ordered the key uncertainties and driving forces, the second step of the analysis is scenario development along two dimensions: the dominance of superintelligence over humans and its impact on population health. The developed scenarios are then substantiated with relevant empirical evidence of current use cases and their respective scenario logic. In the final step, we conduct a strategic foresight analysis to derive policy implications for each scenario, which are outlined in Tables 1-4. For each of the scenarios, one empirical case study is used as a lens to identify key areas for action and outline relevant policy recommendations.

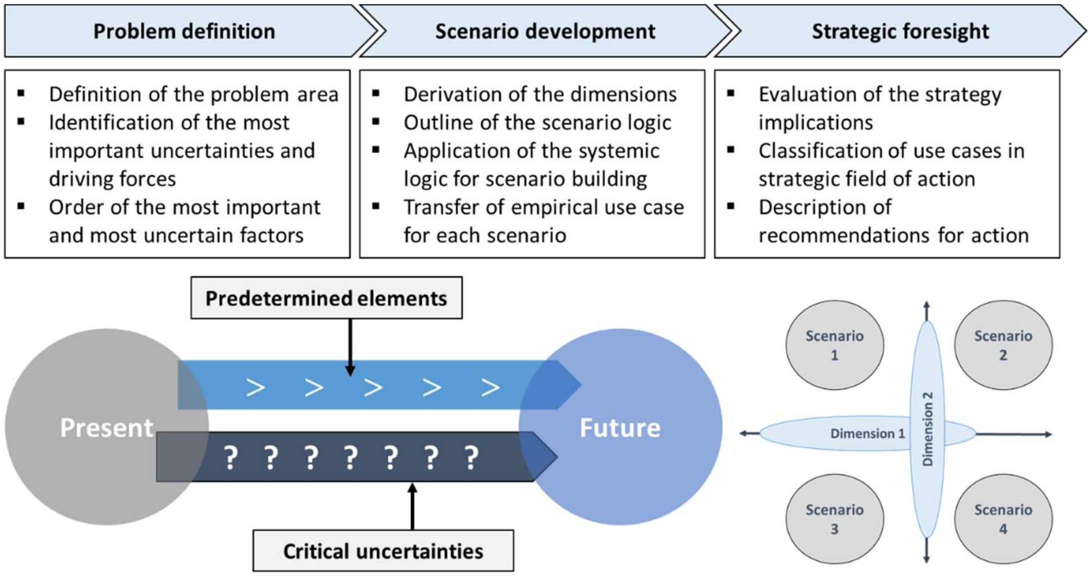

Problem definition
Scenario development
Strategic foresight
▪ Definition of the problem area
▪ Identification of the most important uncertainties and driving forces
▪ Order of the most important and most uncertain factors
▪ Derivation of the dimensions
▪ Outline of the scenario logic
▪ Application of the systemic logic for scenario building
▪ Transfer of empirical use case for each scenario
▪ Evaluation of the strategy implications
▪ Classification of use cases in strategic field of action
▪ Description of recommendations for action
Predetermined elements
Present
> > > > >
? ? ? ? ? ? ?
Future
Critical uncertainties
Scenario 1
Scenario 2
Dimension 1
Dimension 2
Scenario 3
Scenario 4

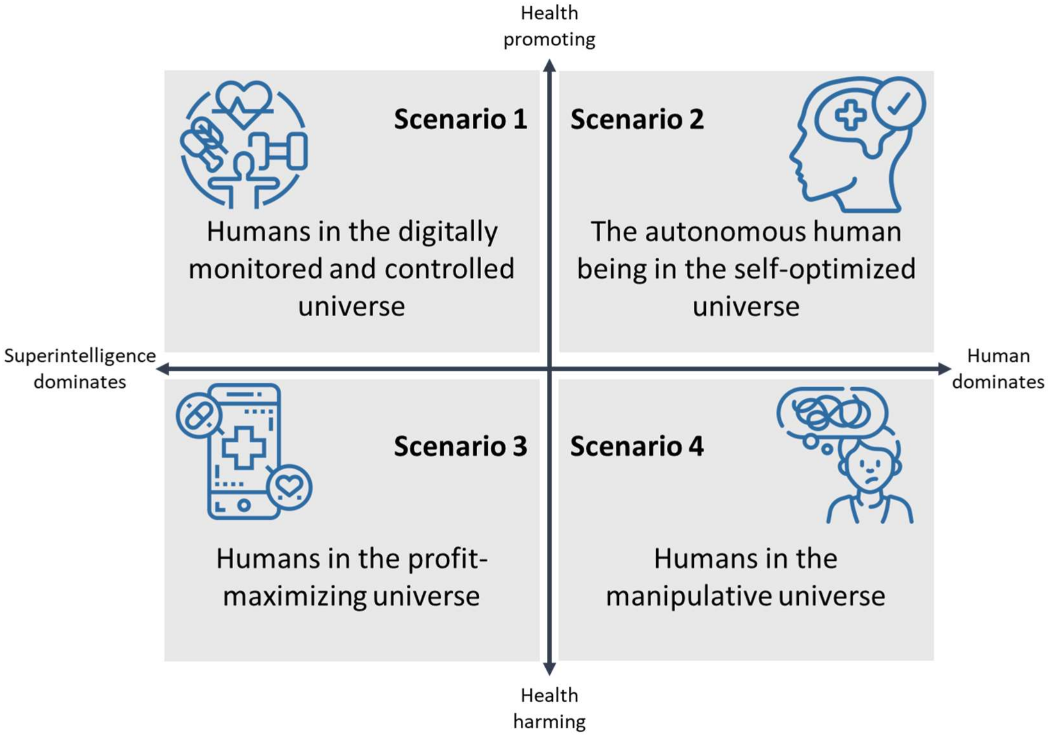
Fig 2: Methodical approach

In building on Reinhart and Greiner's base scenarios, we slightly deviated from the general approach to scenarios based on intuitive logic (Ramirez & Wilkinson, 2014). Rather than explicitly identifying the driving forces, we adapted their model and extracted appropriate descriptive characteristics that supported our dimensions with empirical evidence. We then assigned corresponding predictive attributes for each scenario (Rowe et al., 2017).To ensure the quality of the scenario development process, a panel of subject matter experts tested the internal consistency of the resulting scenarios. For this purpose, a consistency matrix was created to reveal contradictions and present stable relationships (Marthaler et al., 2020). They were discussed within the team of authors and subjected to a plausibility check. The coherent logic of the scenarios with their relation to population health could thus be ensured. Finally, the quality of the developed scenarios was established by triangulating them with empirical findings on current use cases. By means of empirical analogies, scenarios lying far into the future were hence placed in a context of findings available today. Even though these cannot be employed as a clear analogy with their present orientation, they can still serve as metaphors for a critical view of the future (Fischer & Marquardt, 2022).

Our approach allows us to outline scenario narratives with concrete use cases that describe impacts that are already emerging today. This facilitates the identification of opportunities and risks by including predetermined elements from empiricism and critical uncertainties in our considerations. As a result, we were able to derive four possible future development and highlight relevant fields of action (Bradfield et al., 2016; Spaniol & Rowland, 2019).

Results

The four developed scenarios are presented in Figure 3. To clarify the global perspective and relevance for all of humanity, the term ‘universe’ is employed to describe the scenarios. Humans are designated as the supporting object in their relevant universes, reflecting the human-centric component of the scenarios. The first two scenarios, which we consider positive scenarios in the overall context, depict humans in digitally monitored and controlled or self-optimized universes. Both scenarios have a health-promoting effect, and benefit humanity in general, regardless of whether superintelligence dominates or humans take the dominant role with superintelligence support. In contrast, the third scenario depicts humans in a profit-maximizing universe, while the fourth scenario portrays humans in a manipulative universe. These scenarios can be considered negative, as they tend to have a harmful effect on health. Tables 1-4 provide more detailed descriptions of the individual characteristics of each scenario.


Health promoting

Scenario 1

Humans in the digitally monitored and controlled universe

Scenario 2

The autonomous human being in the self-optimized universe

Superintelligence dominates

Human dominates

Scenario 3

Humans in the profit-maximizing universe

Scenario 4

Humans in the manipulative universe

Health harming

Fig 3: Superintelligence and human health

Scenario 1: Humans in the digitally monitored and controlled universe

In this scenario, superintelligence has a dominant position and positively affects human health (cf. Table 1). The superintelligence supports healthy lifestyles, which improves public health and life expectancy. It has access to the latest study results and can derive the best preventive measures or therapy decisions. As a result, health-promoting interventions can be fully integrated into the everyday life of the population (Fozard et al., 2009). Similarly, the superintelligence utilizes behavioral techniques such as nudging to directly and indirectly influence the population in their health behaviors (Felkers et al., 2015). While this approach may lead to a loss of autonomy for individuals, it results in improved health outcomes, creating a trade-off between independent action and decision-making versus health promotion. In always acting rational, the superintelligence can eliminate irrational decisions related to health-damaging behaviors, leading to passive and later active health promotion across the population, enhancing the general state of health, quality of life, and life expectancy (Walorska, 2020).

From a global perspective, the superintelligence can help achieve the United Nations' Sustainable Development Goals, while also reducing a state's healthcare spending by promoting healthy aging and preventing illness across populations (Vinuesa et al., 2020). Despite a potential increase in pension and other social expenditure, healthier individuals can work longer, accumulate more assets, and avoid income or wealth losses due to long absences from illness over their lifetimes. This approach can also lead to the discarding of any behavior that is harmful to health over the long run, as the population is educated to behave in ways that benefit the community. The effect is particularly large for socially disadvantaged groups, as they are disproportionately affected by health-damaging lifestyles (Hosny & Aerts, 2019). From an economic perspective, this effect increases productivity and reduces absenteeism by enabling early detection of potential occupational accidents or work-related illnesses (Badri et al., 2018).

One potential concern with superintelligence is that it may be perceived as paternalistic, as it limits individual’s agency and instead enforces health-conscious behavior in line with societal norms. While this may have long-term benefits, there is a risk that in authoritarian or centralized states, superintelligence may be manipulated to align with the ideas of the ruling class, potentially resulting in negative health effects (Kaplan & Haenlein, 2019). The Chinese social credit system provides an example of such a scenario, where AI is used to monitor and sanction or reward the behavior of the population according to predefined standards, thereby educating the population and influencing their behavior (Creemers, 2018; Yu et al., 2015). Similarly, a superintelligence could implicitly control or influence the overall health behavior of a population, given extensive and ongoing monitoring, and thereby substantially restrict individual’s privacy and autonomy (Roberts et al., 2021).

Table 1: Summary characteristics of the first scenario

Criterion	Description scenario 1
Superintelligence	*dominating*
Health impact	*positive*
Characteristics	*AI supports healthy lifestyle and management* *Consideration of the latest studies (fully automated)* *Expansion of health-promoting measures in the everyday lives of citizens* *Nudging is used to sustainably change health behavior*
Effect	*Loss of autonomy (citizens) in favor of better health* *Passive health promotion of the population as a whole* *Increase in health status and quality of life/expectancy* *Sustainable reduction of health care expenditures*
Opportunities	*Eliminating behavior that is harmful to health (in the long term)* *Educating citizens to adopt optimal health behaviors for the benefit of all* *Particularly large effect on socially disadvantaged groups* *Increasing productivity and reducing absenteeism*
Threats	*Negatively perceived liberal paternalism* *Manipulation risks in authoritarian/centralist forms of government*
Empirical analogy	
Use case	*The Chinese social credit system for behavior management*
Specifics	*AI-supported behavior monitoring and evaluation* *Reward system for socially compliant behavior* *Punishment system for behavior detrimental to society* *Permanent and comprehensive surveillance necessary*
Strategy implications	*Creation of general acceptance of the full monitoring system* *Securing the control mechanisms in the algorithm* *Sharing of power and control (conflicts of interest)*

To establish a system as described above, it is necessary to gain general acceptance of a comprehensive surveillance system. Some argue that surveillance for the benefit of public health is a common good that supersedes individual privacy rights (Fontes et al., 2022). While some parts of population may be willing to allow themselves to be directed by the government, the response to the COVID-19 pandemic has highlighted significant population concerns regarding health surveillance (Liu & Zhao, 2021). One approach policymakers could take is to use the rise of surveillance technologies to counter external risks, such as terrorist attacks, to implement such a system. Still, it would be crucial to incorporate control mechanisms as safeguards to monitor the algorithm and provide transparency to public authorities. Additionally, there must be a clear division of power and control. If a superintelligence decides and influences the lives and health of millions of people, the balance of power and adequate independent monitoring must be ensured.

Scenario 2: The autonomous human in a self-optimized universe

In this scenario, humans retain their superior position over the superintelligence and the health impact is positive (cf. Table 2). Humans deploy the superintelligence as a tool to optimize their health while making information and decision-making processes fully transparent. Assuming a wide availability of medical innovations and a deep understanding of health-affecting behaviors, the ultimate goal is to achieve the best possible health and maximize well-being (Haselager & Mecacci, 2020). To this end, the population actively engages in targeted health promotion, resulting in overall improvements in health and increased life expectancy. Moreover, people's adherence to health behaviors also increases, for example in AI assisted medication dispensary (Shaban-Nejad et al., 2018) . Through improved health, the participation opportunities of the population are also enhanced.

The main benefits of this scenario include promoting autonomy and freedom of choice, reducing health disparities, and increasing productivity. These factors can lead to greater satisfaction and unlock significant economic potential. Unlike in scenario one, individuals here enjoy greater personal freedom and can fully achieve their capabilities and developmental potential. However, despite the support of superintelligence, achieving optimal health outcomes requires a higher degree of health literacy. In addition, social inequalities could persist (Dunn & Hazzard, 2019), particularly if access to authoritative medical innovations is limited or if individuals persist in engaging in irrational behaviors contrary to the recommendations of superintelligence.

Empirically, this scenario is exemplified by the Technological Singularity of Silicon Valley (Solez et al., 2013) which is characterized by a high degree of openness to technology and innovation. It follows a utopian paradigm and is not commonly found in practice. The approach is built upon the potentials of exponential medicine (Nabipour & Assadi, 2016), which is considered to offer quasi-infinite technological possibilities, potentially leading to a significant increase in life expectancy. Additionally, the use of resources is optimally controlled with the help of superintelligence, aiming to achieve healthy living for all (Popa, 2014). In principle, these opportunities should be available to the entire population, but there is a risk of exclusivity for privileged groups leading to both national and global inequalities in healthcare access. Strategic implications for realizing this scenario include investing in health education and independent information to complement support for superintelligence, increasing public trust. State institutions should ensure universal equitable access to prevent social division. Additionally, a supervisory authority is crucial to prevent a market-dominating position and control the mode of operation of superintelligence, especially if many private-sector companies are involved, which could result in significant conflicts of interest.

Table 2: Summary characteristics of the second scenario

Criterion	Description scenario 2
Superintelligence	*not dominating*
Health impact	*positive*
Characteristics	*Citizens use AI systems to self-optimize their health* *Transparent information and decision-making processes* *Wide availability of medical innovation and its mode of action*
Effect	*Focus on health and well-being as a new lifestyle* *Active health promotion of the entire population* *Increasing health status and quality/expectancy of life* *Increase compliance and paticipation opportunities*
Opportunities	*Promotion of autonomy and freedom of choice* *Reduction of health inequalities possible* *Increase productivity and reduce absence from work*
Threats	*High health literacy required* *Social inequality could increase (lack of access)*
Empirical analogy	
Use case	*Technological Singularity in Silicon Valley*
Specifics	*Exploiting the potential of exponential medicine (incl. technologies)* *Strong prolongation of life expectancy* *Optimal use of resources for a healthy life* *Long-term access for all people (risk of exclusivity)*
Strategy implications	*Investing in health education and independent information* *Ensure universal equitable access* *Central "good" supervisory authority necessary (market control)*

Scenario 3: Humans in a profit-maximizing universe

In this scenario, superintelligence holds a dominant position, and its impact on health tends to be negative (cf. Table 3). While this is not a guaranteed outcome, it is highly likely given the constellation of this scenario. AI-driven utility maximization is the primary objective, with companies employing superintelligence to maximize profits (Leggett, 2021). All other goals are subordinate to this primary objective. Therefore, superintelligence will optimize a company's profits, even if it results in harm to health. In the current capitalist system, groundbreaking technological innovations often stem from fundamental research funded by governments that later get commercialized by companies. There is no evidence to suggest that this paradigm would change in the case of superintelligence of the underlying systems. Another possibility is that a super intelligent AI might prioritize economic factors for the greater good of the species, even if it comes at the expense of human health. However, it remains unknown which priority goals governments and companies will ultimately choose.

The consequences of the pursuit of profit maximization by companies through superintelligence can be severe, leading to negative impacts on population health and well-being. As companies prioritize their profits over public health interests, they may not take adequate measures to address the negative effects of their products or services, leading to the promotion of disease patterns or the emergence of new health risks (Hou et al., 2019). For example, social media consumption can lead to mental health problems with long-term consequences, contributing to an overall deterioration of population health and increased healthcare costs. Moreover, tech companies could develop dominant market positions vis-à-vis state institutions, as their superintelligences could not be controlled or regulated from the outside (Kaplan & Haenlein, 2020). This can lead to social division and diminished population health, which in turn can have significant economic impacts, including reduced productivity and increasing healthcare expenditures.

However, there are opportunities in this scenario for state institutions to cooperate with the digital industry in using the positive elements of superintelligence profitably. Companies may have to forego some profits, but in return, they would benefit from other privileges offered to them by the state. Similar to the second scenario, the data available from large technology companies that use superintelligence, could also be used for public health promotion. The state could act as a customer of these companies for this purpose. However, there are numerous risks associated with this negative scenario. Companies could exploit their position of power, as commercially active companies are obligated to their shareholders and not to the public. An ethical dilemma could arise regarding how superintelligence is used, with profit maximization taking precedence over all other interests, which could jeopardize social cohesion. Additionally, there are issues around the regulation of superintelligence as private companies likely have strong information asymmetry towards policymakers, for example with regards to the source code and primary data.

Empirically, this scenario can be illustrated by the profit maximization paradigm of Facebook (Frost & Rickwood, 2017). While recently changing its name to Meta, internal leaks disclosed a range of potentially harmful business practices that favored higher profits to the detriment of public health. This included a range of issues from the Facebook AI encouraging user's excessive social media consumption to the dissemination of fake news (Terrasse et al., 2019; Walter et al., 2021). These leaks reveal the real possibility of unethical corporate actions in the use of AI and the associated lack of social responsibility. Furthermore, the case illustrates how the lack of transparency in these activities means that politicians and the public have no means of monitoring the activities of these companies (Roland, 2018). The resulting strategic implications are the regulation of monopoly-like structures, the creation of universally applicable compliance rules for the technology companies, and the establishment of powerful supervisory bodies. Ethical standards must be defined, and non-compliance must be sanctioned. Additionally, information asymmetries between private and public sector actors must be significantly reduced to enable adequate superintelligence regulation.

Table 3: Summary characteristics of the third scenario

Criterion	Description scenario 3
Superintelligence	*dominating*
Health impact	*negative*
Characteristics	*AI-driven utility maximization as the ultimate goal* *Benefit maximization in favor of one entity, at the expense of the population* *Collateral damage is accepted* *Overpowering positions vis-à-vis state institutions*
Effect	*Ethical and moral fault lines* *Promotion of disease patterns, emergence of new "widespread" diseases* *Deteriorating health status of the population as a whole* *Exploding health care expenditures*
Opportunities	*Government and digital industry collaborations (focus on positive potentials)* *Use of data/algorithms for public health promotion*
Threats	*Exploitation of power positions on the part of Big Tech* *Threat to social cohesion* *Lack of know-how/expertise among regulators/legislators* *Decreasing productivity and increasing downtime*
Empirical analogy	
Use case	*Profit maximization paradigm of the company Facebook/Meta*
Specifics	*Excessive increase in the use of social media (Facebook, Instagram, etc.)* *Exploitation of profit potential as top corporate maxim* *Lack of social responsibility* *Opacity and lack of transparency towards politics and the public*
Strategy implications	*Regulation of monopoly-like structures* *Creation of general digital compliance for big tech* *Establishment of powerful supervisory bodies*

Scenario 4: Humans in the manipulative universe.

In this scenario, humans have a dominant position and use superintelligence in a purposefully manipulative manner, creating a negative health effect and in some cases actively promoting it (cf. Table 4). AI is employed to serve individual interests, such as the dissemination of fake news or the manipulation of public discourse through bots (Vafeiadis et al., 2019). Such practices occur at both the state and institutional level and have significant strategic implications in global competition, as they shape narratives and sway public opinion. This often results in collateral damage, which is accepted or even encouraged (Monsees, 2020).

The effects are reflected in a high level of information insecurity. The population loses trust in state institutions or questions recognized scientific methods, including in diagnostics and therapy (Mesquita et al., 2020). As a result, the health status of the population deteriorates because of unclaimed health services, avoidance of preventive services, and a distrust of government programs, such as vaccination campaigns (Carrieri et al., 2019). In the medium and long term, the negative change in health status leads to higher health risks and higher health expenditures to address the consequential damage.

Opportunities in this scenario would be particularly evident in supranational cooperation among states, which could strengthen overall health diplomacy (Fazal, 2020). In addition, greater efforts could be made to provide adequate information, improve transparency, and increase responsible actors' expertise in dealing with manipulative AI. Nonetheless, there are numerous risks in this negative scenario. Trust in state institutions is damaged, with long-term repercussions on social cohesion (Braunschweig & Ghallab, 2021). The responsible control bodies are overwhelmed by the use of technology and cannot respond appropriately to the activities of AI. Deteriorated health results in lower productivity and increased absenteeism in the population. In this scenario, these risks are not just passively accepted, but can be part of (geo-)political strategies to improve one's power or competitive position.

Empirically this scenario can be illustrated by the "information war" regarding the COVID-19 vaccine (Carrion-Alvarez & Tijerina-Salina, 2020). With the help of fake news and bots, targeted disinformation is disseminated on a large scale to undermine trust in the Corona vaccine by propagating serious vaccine harms or spreading conspiracy myths (Catalan-Matamoros & Elías, 2020). As a result, willingness to receive the vaccine is declining, and two opposing blocs have emerged in the population with vaccination supporters and opponents at the margins (Borkowska & Laurence, 2021). This leads to polarization and division in society, which can also have significant effects on the stability of the political system (Jahng, 2021). As a result, a long-term skepticism toward vaccination programs and conventional medicine may emerge.

Table 4: Summary characteristics of the fourth scenario

Criterion	Description scenario 4
Superintelligence	*not dominating*
Health impact	*negative*
Characteristics	*Use of AI to enforce individual interests (fake news, bots, etc.)* *High strategic implication in global competition* *Struggle for narratives and public opinion sovereignty* *(High) collateral damage is accepted*
Effect	*Information uncertainty* *Loss of confidence in recognized diagnostic and therapeutic procedures* *Deteriorating health status of the population as a whole* *Rising health care expenditure/risks*
Opportunities	*Supranational cooperation of the states (health diplomacy)* *Compulsion to improve the quality of information and know-how appropriation*
Threats	*Long-term damage to trust in state institutions* *Threat to social cohesion* *Technical overload of the control authorities* *Decreasing productivity and increasing downtimes*
Empirical analogy	
Use case	*The "Information War" on vaccination against the Corona virus*
Specifics	*Dissemination of Fake News about vaccination harms and conspiracy myths* *Formation of opposing blocs (pro/contra vaccination)* *Social polarization and division* *Long-term skepticism in vaccination programs and conventional medicine*
Strategy implications	*Creation of "secure" and trustworthy information channels* *Transparent communication and decision-making* *Regulation of bot-prone systems (including social media)*

Strategic implications in this scenario include the creation of secure and trustworthy information channels, transparent communication and decision-making and adequate regulation of AI systems that affect population health. Ideally, independent nongovernmental institutions, with a focus on avoiding potential conflicts of interest, should provide information to the public. Social media networks, which are susceptible to bot manipulation, should be subject to community regulation at the international level or, if necessary, their use severely restricted if companies refuse to comply with established standards. It is crucial to take a proactive approach to address the risks of manipulative AI in order to maintain public trust in health institutions and ensure the well-being of the population.

Future perspective of the scenarios

In our four scenarios, we described the potential opportunities and risks for population health posed by superintelligence. In order to translate the strategic implications derived from our scenarios into policy frameworks for global population health, it may be useful to draw an analogy to the existential threat posed by climate change (Schuppert, 2011). Like the threat of climate change, the risks associated with the use of superintelligence require a collaborative, consensus-oriented approach to problem-solving. At the global level, institutions such as the United Nations exist to combat climate change, and national representatives negotiate compromises to solve problems, which are then translated into national legislation (Gao et al., 2017). For example, in Germany, Article 20a was enshrined in the constitution, giving people an enforceable basic right to secure the future for future generations (Griefahn, 1999). A similar construct would be possible to safeguard global population health by actively reducing the risks described at a higher level. At the World Health Organization (WHO) level, nations could agree on a basic set of rules for the use of algorithms by a future superintelligence. Universal algorithm laws could be formulated as a basis for this set of rules. These universal laws could be derived from Asimov's robot laws (Clarke, 1994). Asimov's remarks were the first to lay an ethical foundation in the interaction of robots and humans. Asimov's reasoning is based on metaphors from science fiction literature, which made important ethical contributions to the knowledge of futurology (Blackford, 2017). Here, the protection of human life is set as the highest premise, which would also be a purposeful analogy for superintelligence algorithms (Nagler et al., 2019).

One potential next step in safeguarding population health against superintelligence is to enshrine algorithm laws in the constitution, which would provide better options for legislative and jurisdictional action at all levels. A general regulation through ordinances, for example at the level of the European Union for its member states, is also a conceivable approach. There are already attempts to do so in parliamentary processes (Robles Carrillo, 2020; Schneeberger et al., 2020). However, the precautionary principle (Calliess, 2013) that is preferred in EU legislation is challenging to apply to digital services. Unlike physical products that require proof of the exclusion of harmful effects on health before being put into circulation, practical and methodological problems make it difficult to apply the same principle to algorithms (Kim, 2019).

Currently, a new EU digital legislation is underway to better regulate the digital economy. The EU Commission is pioneering ex-ante regulation in local legislation (Georgieva, 2021) to address this challenge. However, these draft laws for regulating algorithms used by private companies also have weaknesses as they are not comprehensively enforceable or sanctionable for violations. Therefore, adopting the risk principle as applied in United States (Peuker, 2014) could serve as a model for the regulation of algorithms. This principle sets preventive incentives to exclude possible consequential damages in advance due to a risk of lawsuits. Companies would design and program the algorithms of superintelligences in such a way that no adverse health effects occur because of them. Otherwise, national legal standards, such as class actions, offer the possibility of holding them accountable. This ex-ante regulation approach would set appropriate incentives for companies and states to minimize health harms at the expense of population health through AI. As a result, existential risks from the scenarios described could be significantly reduced or eliminated.

Discussion

In this paper, we have derived and analyzed scenarios how superintelligence could have an impact on population health and pose an existential risk to the future of humanity. This is important not only from the perspective of intergenerational justice but also because economic prosperity, health and social systems, and social cohesion are all dependent on population health. As a precautionary measure, we propose enshrining algorithm laws in the constitution to establish positive incentives for companies and institutions to use future superintelligences in a responsible manner.

To derive these scenarios, we used an appropriate method for anticipating future developments and their consequences. We adapted an existing scenario model to include the relevant dimensions of population health and validated them through extensive discussions and empirical evidence. Our scenarios are not exhaustive and should be interpreted with caution since future developments are inherently uncertain. However, they provide a basis for further discussion and empirical research.

The accuracy of the four scenarios is supported by existing facts, allowing us to develop narratives for each scenario that describe their impacts, opportunities, and risks, as well as derive strategic implications. Future avenues for research could further validate and refine our scenarios and provide a more comprehensive understanding of potential future developments. While none of our scenarios will occur in their described pure form, they do offer insights into the possible risks and opportunities associated with superintelligences and their impact on population health. Our study is a desk research work that does not involve participatory empirical data collection. We chose this type of study in order to first adopt an exploratory approach, as there are no theoretical frameworks in this field that explicitly deal with public health. In our study, a positive and negative classification of scenarios emerges. We do not give this classification normatively, but it results from the effect of a potentially deteriorating health of the population. We point out that this classification according to the scenario effect is an anticipation assumed by the authors. The authors assume that there is a consensus in the scientific community that an adverse health effect is always negatively associated.

One limitation of the developed scenarios is the present orientation of their illustrative use cases. Consequently, these serve to illustrate current trends, which are intended to show a scenario path to bring the imagination about possible future closer (Suddendorf & Redshaw, 2013). This speculative nature is inherent in scenario analysis, but it is a useful tool for anticipating different perspectives and to question, accompany and renew them in a learning process of advancing developments (Rhisiart et al., 2015). While the scenarios presented use catchy labels to highlight their narrative, the factors selected to determine their logic follow a consistent and plausible approach.

At the outset, we acknowledged that we cannot predict how a superintelligence will behave or what features it will possess. However, there are currently existing functionalities that we can examine, such as the complex question processing ability of ChatGPT from OpenAI (Mijwil et al., 2023). This provides an indication of the potential direction of interaction between a strong AI and individuals seeking health information. It should be noted, however, that the transparency of these chatbots is insufficient and general acceptance of the information they provide is unknown. Additionally, there are many studies that demonstrate the superior diagnostic quality of strong AI in some domains. Nevertheless, these studies are not unified under a superintelligence in a broader sense, and it remains uncertain whether and in what form they will be integrated into a uniform data platform for the use of superintelligence.

Conclusion

In this paper, we explored the potential impact of superintelligence on population health and the associated existential risks. To do this, we used the status quo of global population health as a starting point and developed four scenarios using the intuitive logic method to explore their potential outcomes. Scenario one depicts humans in a digitally monitored and controlled universe. Empirical evidence is provided by the Chinese social credit system, which is used as an analogy for fully controlled health behavior by a superintelligence. Scenario two describes autonomous humans in a self-optimized universe. Here, a superintelligence is specifically deployed by humans to optimize population health in the spirit of Technological Singularity and to exploit the full innovation potential of medicine. The third scenario shows humans in a profit-maximizing universe. Here, economic interests always take precedence over those of population health, which can result in significant negative health effects, as the case study of Facebook shows. The fourth scenario illustrates humans in the manipulative universe. Here, targeted instruments are used to enforce, among other things, geostrategic interests that can have strong negative effects on population health. In the negatively interpreted scenarios, superintelligence does not serve to increase health status but to achieve self-interest that opposes population health or accepts collateral damage. In this paper, we provide an analytically derived description of future scenarios, make recommendations for action, and describe a framework for regulating superintelligence to minimize existential risks to population health.

Overall, we recommend a policy of close coordination between all stakeholders involved in the development, use, and regulation of a superintelligence. Already during the development of this technology, legislators and civil society must address these complex ethical issues and engage in an open discourse. Similarly, it is crucial that policymakers build expertise in this field to limit information asymmetries to private sector stakeholders. Technologies with a high impact on population health and social cohesion must be subject to ethical standards when commercialized, which are to be defined on the global, regional, and national level.

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database, 2020*.

Badri, A., Boudreau-Trudel, B., & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern? *Safety science, 109*, 403-411.

Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2020). The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health, 74*(11), 964-968.

Batin, M., Turchin, A., Sergey, M., Zhila, A., & Denkenberger, D. (2017). Artificial intelligence in life extension: from deep learning to superintelligence. *Informatica, 41*(4).

Blackford, R. (2017). *Science fiction and the moral imagination: visions, minds, ethics*. Springer.

Borkowska, M., & Laurence, J. (2021). Coming together or coming apart? Changes in social cohesion during the Covid-19 pandemic in England. *European Societies, 23*(sup1), S618-S636.

Bostrom, N. (1998). How long before superintelligence? *International Journal of Futures Studies, 2*.

Bradfield, R., Derbyshire, J., & Wright, G. (2016). The critical role of history in scenario thinking: Augmenting causal analysis within the intuitive logics scenario development methodology. *Futures, 77*, 56-66.

Bradfield, R. M. (2008). Cognitive barriers in the scenario development process. *Advances in Developing Human Resources, 10*(2), 198-215.

Braunschweig, B., & Ghallab, M. (2021). *Reflections on artificial intelligence for humanity*. Springer.

Breyer, F., & Felder, S. (2006). Life expectancy and health care expenditures: a new calculation for Germany using the costs of dying. *Health policy, 75*(2), 178-186.

Brundage, M. (2015, 2015/09/01/). Taking superintelligence seriously: Superintelligence: Paths, dangers, strategies by Nick Bostrom (Oxford University Press, 2014). *Futures, 72*, 32-35. https://doi.org/https://doi.org/10.1016/j.futures.2015.07.009

Calliess, C. (2013). Vorsorgeprinzip. In *Handbuch Technikethik* (pp. 390-394). Springer.

Cardis, E., & Hatch, M. (2011). The Chernobyl accident—an epidemiological perspective. *Clinical Oncology, 23*(4), 251-260.

Carrieri, V., Madio, L., & Principe, F. (2019). Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health economics, 28*(11), 1377-1382.

Carrion-Alvarez, D., & Tijerina-Salina, P. X. (2020). Fake news in COVID-19: A perspective. *Health promotion perspectives, 10*(4), 290.

Catalan-Matamoros, D., & Elías, C. (2020). Vaccine hesitancy in the age of coronavirus and fake news: analysis of journalistic sources in the Spanish quality press. *International journal of environmental research and public health, 17*(21), 8136.

Clarke, R. (1994). Asimov's laws of robotics: Implications for information technology. 2. *Computer, 27*(1), 57-66.

Creemers, R. (2018). China's Social Credit System: an evolving practice of control. *Available at SSRN 3175792*.

Dunn, P., & Hazzard, E. (2019). Technology approaches to digital health literacy. *International journal of cardiology, 293*, 294-296.

Fazal, T. M. (2020). Health diplomacy in pandemical times. *International Organization, 74*(S1), E78-E97.

Felkers, I., Maclean, M., & Mulder, E. (2015). Homo ludens in the Twenty-first Century. *Philosophical Perspectives on Play*, 124-135.

Fischer, N., & Marquardt, K. (2022). Playing with Metaphors. Connecting Experiential Futures and Critical Futures Studies. *Journal of Futures Studies, 27*(No.1).

Fontes, C., Hohma, E., Corrigan, C. C., & Lütge, C. (2022, 2022/11/01/). AI-powered public surveillance systems: why we (might) need them and how we want them. *Technology in Society, 71*, 102137. https://doi.org/https://doi.org/10.1016/j.techsoc.2022.102137

Fozard, J. L., Bouma, H., Franco, A., & Van Bronswijk, J. (2009). Homo ludens: Adult creativity and quality of life. *Gerontechnology, 8*(4), 187-196.

Frost, R. L., & Rickwood, D. J. (2017). A systematic review of the mental health outcomes associated with Facebook use. *Computers in Human Behavior, 76*, 576-600.

Gao, Y., Gao, X., & Zhang, X. (2017). The 2 C global temperature target and the evolution of the long-term goal of addressing climate change—from the United Nations framework convention on climate change to the Paris agreement. *Engineering, 3*(2), 272-278.

Georgieva, Z. (2021). The Digital Markets Act Proposal of the European Commission: Ex-ante Regulation, Infused with Competition Principles. *European Papers-A Journal on Law and Integration, 2021*(1), 25-28.

Giacomelli, L., Perrotta, A., Scandone, R., & Scarpati, C. (2003). The eruption of Vesuvius of 79 AD and its impact on human environment in Pompeii. *Episodes-Newsmagazine of the International Union of Geological Sciences, 26*(3), 235-238.

Griefahn, M. (1999). Der Schutz der Umwelt als Menschenrecht? In *Menschenrechte und Bürgergesellschaft in Deutschland* (pp. 159-164). Springer.

Haselager, P., & Mecacci, G. (2020). Superethics instead of superintelligence: know thyself, and apply science accordingly. *AJOB neuroscience, 11*(2), 113-119.

Hosny, A., & Aerts, H. J. (2019). Artificial intelligence for global health. *Science, 366*(6468), 955-956.

Hou, Y., Xiong, D., Jiang, T., Song, L., & Wang, Q. (2019). Social media addiction: Its impact, mediation, and intervention. *Cyberpsychology: Journal of psychosocial research on cyberspace, 13*(1).

Huss, W. R., & Honton, E. J. (1987). Scenario planning—what style should you use? *Long range planning, 20*(4), 21-29.

Huston, P., Edge, V. L., & Bernier, E. (2019, Oct 3). Reaping the benefits of Open Data in public health. *Can Commun Dis Rep, 45*(11), 252-256. https://doi.org/10.14745/ccdr.v45i10a01

Ioannou, A., & Tussyadiah, I. (2021, 2021/11/01/). Privacy and surveillance attitudes during health crises: Acceptance of surveillance and privacy protection behaviours. *Technology in Society, 67*, 101774. https://doi.org/https://doi.org/10.1016/j.techsoc.2021.101774

Jahng, M. R. (2021). Is fake news the new social media crisis? Examining the public evaluation of crisis management for corporate organizations targeted in fake news. *International Journal of Strategic Communication, 15*(1), 18-36.

Johnson, I., Hansen, A., & Bi, P. (2018). The challenges of implementing an integrated One Health surveillance system in Australia. *Zoonoses and Public Health, 65*(1), e229-e236. https://doi.org/https://doi.org/10.1111/zph.12433

Kachanoff, F. J., Bigman, Y. E., Kapsaskis, K., & Gray, K. (2021). Measuring Realistic and Symbolic Threats of COVID-19 and Their Unique Impacts on Well-Being and Adherence to Public Health Behaviors. *Social Psychological and Personality Science, 12*(5), 603-616. https://doi.org/10.1177/1948550620931634

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25.

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons, 63*(1), 37-50.

Kim, M. (2019). *Roboterrecht in der modernen Gesellschaft: Vorschläge zur Gesetzgebung und Reform*. Logos Verlag Berlin GmbH.

Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. (2021, 2021/02/01). A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine, 27*(2), 225-228. https://doi.org/10.1038/s41591-020-1124-9

Leggett, D. (2021). Feeding the Beast: Superintelligence, Corporate Capitalism and the End of Humanity. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,

Liu, J., & Zhao, H. (2021, 2021/11/01/). Privacy lost: Appropriating surveillance technology in China's fight against COVID-19. *Business Horizons, 64*(6), 743-756. https://doi.org/https://doi.org/10.1016/j.bushor.2021.07.004

Marthaler, F., Gesk, J. W., Siebe, A., & Albers, A. (2020, 2020/01/01/). An explorative approach to deriving future scenarios: A first comparison of the consistency matrix-based and the catalog-based approach to generating future scenarios. *Procedia CIRP, 91*, 883-892. https://doi.org/https://doi.org/10.1016/j.procir.2020.02.245

Mercer, C., & Trothen, T. J. (2021). Superintelligence: Bringing on the Singularity. In *Religion and the Technological Future* (pp. 181-204). Springer.

Meskó, B., Hetényi, G., & Győrffy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC health services research, 18*(1), 1-4.

Mijwil, M., Mohammad, A., & Ahmed Hussein, A. (2023, 02/01). ChatGPT: Exploring the Role of Cybersecurity in the Protection of Medical Information. *Mesopotamian Journal of CyberSecurity, 2023*, 18-21. https://doi.org/10.58496/MJCS/2023/004

Monsees, L. (2020). 'A war against truth'-understanding the fake news controversy. *Critical Studies on Security, 8*(2), 116-129.

Nabipour, I., & Assadi, M. (2016). The technological singularity and exponential medicine. *ISMJ, 18*(6), 1287-1298.

Nagler, J., Hoven, J. v. d., & Helbing, D. (2019). An extension of asimov's robotics laws. In *Towards Digital Enlightenment* (pp. 41-46). Springer.

Peuker, B. (2014). Der Streit um die Agrar-Gentechnik. In *Der Streit um die Agrar-Gentechnik*. transcript-Verlag.

Popa, S. (2014). Exponential Medicine Conference. *San Diego, California, USA*.

Ramirez, R., & Wilkinson, A. (2014). Rethinking the 2× 2 scenario method: Grid or frames? *Technological Forecasting and Social Change, 86*, 254-264.

Reinhart, J., & Greiner, C. (2019). Künstliche Intelligenz–eine Einführung. *Grundlagen, Anwendungsbeispiele und Umsetzungsstrategien für Unternehmen*.

Rhisiart, M., Miller, R., & Brooks, S. (2015, 2015/12/01/). Learning to use the future: developing foresight capabilities through scenario processes. *Technological Forecasting and Social Change, 101*, 124-133. https://doi.org/https://doi.org/10.1016/j.techfore.2014.10.015

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & society, 36*(1), 59-77.

Robles Carrillo, M. (2020, 2020/07/01/). Artificial intelligence: From ethics to law. *Telecommunications Policy, 44*(6), 101937. https://doi.org/https://doi.org/10.1016/j.telpol.2020.101937

Roland, D. (2018). Social media, health policy, and knowledge translation. *Journal of the American College of Radiology, 15*(1), 149-152.

Rowe, E., Wright, G., & Derbyshire, J. (2017, 2017/12/01/). Enhancing horizon scanning by utilizing pre-developed scenarios: Analysis of current practice and specification of a process improvement to aid the identification of important 'weak signals'. *Technological Forecasting and Social Change, 125*, 224-235. https://doi.org/https://doi.org/10.1016/j.techfore.2017.08.001

Schneeberger, D., Stöger, K., & Holzinger, A. (2020, 2020//). The European Legal Framework for Medical AI. Machine Learning and Knowledge Extraction, Cham.

Schoemaker, P. J. (1995). Scenario planning: a tool for strategic thinking. *Sloan management review, 36*(2), 25-50.

Schoemaker, P. J. (2004). Forecasting and scenario planning: the challenges of uncertainty and complexity. *Blackwell handbook of judgment and decision making*, 274-296.

Schoemaker, P. J., & van der Heijden, C. A. (1992). Integrating scenarios into strategic planning at Royal Dutch/Shell. *Planning Review*.

Schuppert, F. (2011). Climate change mitigation and intergenerational justice. *Environmental Politics, 20*(3), 303-321.

Solez, K., Bernier, A., Crichton, J., Graves, H., Kuttikat, P., Lockwood, R., Marovitz, W. F., Monroe, D., Pallen, M., & Pandya, S. (2013). Bridging the gap between the technological singularity and medicine: Highlighting a course on technology and the future of medicine. *Global Journal of Health Science, 5*(6), 112.

Sotala, K. (2017). How feasible is the rapid development of artificial superintelligence? *Physica Scripta, 92*(11), 113001.

Spaniol, M. J., & Rowland, N. J. (2019). Defining scenario. *Futures & Foresight Science, 1*(1), e3.

Steptoe, A., Deaton, A., & Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The lancet, 385*(9968), 640-648.

Suddendorf, T., & Redshaw, J. (2013). The development of mental scenario building and episodic foresight. *Annals of the New York Academy of Sciences, 1296*(1), 135-153. https://doi.org/https://doi.org/10.1111/nyas.12189

Terrasse, M., Gorin, M., & Sisti, D. (2019). Social media, e-health, and medical ethics. *Hastings Center Report, 49*(1), 24-33.

Tol, R. S. (2020). The economic impacts of climate change. *Review of Environmental Economics and Policy*.

Torres, P. (2019). Existential risks: a philosophical analysis. *Inquiry*, 1-26. https://doi.org/10.1080/0020174X.2019.1658626

Tortorella, G. L., Fogliatto, F. S., Mac Cawley Vergara, A., Vassolo, R., & Sawhney, R. (2020). Healthcare 4.0: trends, challenges and research directions. *Production Planning & Control, 31*(15), 1245-1260.

Vafeiadis, M., Bortree, D. S., Buckley, C., Diddi, P., & Xiao, A. (2019). Refuting fake news on social media: nonprofits, crisis response strategies and issue involvement. *Journal of Product & Brand Management*.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications, 11*(1), 1-10.

Walorska, A. M. (2020). The Algorithmic Society. In *Redesigning Organizations* (pp. 149-160). Springer.

Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2021). Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication, 36*(13), 1776-1784.

Werther, G. F. A. (2013). When black swans aren't: on better recognition, assessment, and forecasting of large scale, large impact, and rare event change. *Risk Management and Insurance Review, 16*(1), 1-23.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A. G., de Souza Dias, B. F., Ezeh, A., Frumkin, H., Gong, P., & Head, P. (2015). Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. *The lancet, 386*(10007), 1973-2028.

Winch, G. M., & Maytorena-Sanchez, E. (2011). Managing risk and uncertainty on projects: A cognitive approach. In *The Oxford handbook of project management* (pp. 345-364). Oxford University Press.

Wright, G., Bradfield, R., & Cairns, G. (2013). Does the intuitive logics method–and its recent enhancements–produce "effective" scenarios? *Technological Forecasting and Social Change, 80*(4), 631-642.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting, 25*(4), 813-825.

Yu, L., Li, X., Tang, L., Zhang, Z., & Kou, G. (2015). Social credit: a comprehensive literature review. *Financial Innovation, 1*(1), 1-18.