

Article

Chances and Risks of Artificial Intelligence—A Concept of Developing and Exploiting Machine Intelligence for Future Societies

Reinhard Hutter ^{1,*} and Marcus Hutter ²¹ Centre for European Security Strategies, 81545 Munich, Germany² Research School of Computer Science (RSCS), Australian National University, Canberra, ACT 2600, Australia; marcus.hutter@gmx.net

* Correspondence: Reinhard.hutter@gmx.de or hutter@cess-net.eu; Tel.: +49-89-69377147

Abstract: Artificial Intelligence (AI): Boon or Bane for societies? AI technologies and solutions—as most revolutionary technologies have done in the past—offer negative implications on the one hand and considerable positive potential on the other. Avoiding the former and fostering the latter will require substantial investments in future societal concepts, research and development, and control of AI-based solutions in AI security while avoiding abuse. Preparation for the future role of AI in societies should strive towards the implementation of related methods and tools for risk management, models of complementary human–machine cooperation, strategies for the optimization of production and administration, and innovative concepts for the distribution of the economic value created. Two extreme possible “end states” of AI impact (if there is ever an end state) that are being discussed at present may manifest as (a) uncontrolled substitution by AI of major aspects of production, services, and administrative and decision-making processes, leading to unprecedented risks such as high unemployment, and devaluation and the underpayment of people in paid work, resulting in inequality in the distribution of wealth and employment, diminishing social peace, social cohesion, solidarity, security, etc., or, on the contrary, (b) the freeing of people from routine labor through increased automation in production, administration and services, and changing the constitution of politics and societies into constituencies with high ethical standards, personal self-determination, and the general dominance of humane principles, as opposed to pure materialism. Any mix of these two extremes could develop, and these combinations may vary among different societies and political systems.

Keywords: artificial and human intelligence; security; risks and risk management; quality of life; common welfare; socio-political assessment



Citation: Hutter, R.; Hutter, M. Chances and Risks of Artificial Intelligence—A Concept of Developing and Exploiting Machine Intelligence for Future Societies. *Appl. Syst. Innov.* **2021**, *4*, 37. <https://doi.org/10.3390/asi4020037>

Academic Editor: Andrzej Białas

Received: 3 April 2021

Accepted: 12 May 2021

Published: 2 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI) and all its theoretical, technical and operational ramifications offer a vast and rapidly developing potential to serve humankind. The authors were motivated by realizing that most discussions of AI stress either technical matters or the possible negative impacts of AI's expansion into society. The ideas presented here deviate to from today's mainstream public discussions regarding the future role of AI. Although they are a bit revolutionary, or at least unusual, the concepts presented here promise a high payoff for societies derived from AI, while, at the same time, avoiding quite a high number of risks of and controversies. Figure 1 illustrates that, sooner or later, all relevant societal spheres will be impacted by AI-based solutions, and that, in turn, the path of AI development will require guidance and control. This will require an unprecedentedly concerted effort among all societal forces with respect to the preparation, introduction, operation and maintenance of AI solutions, and with respect to the principles that should guide their development and use.

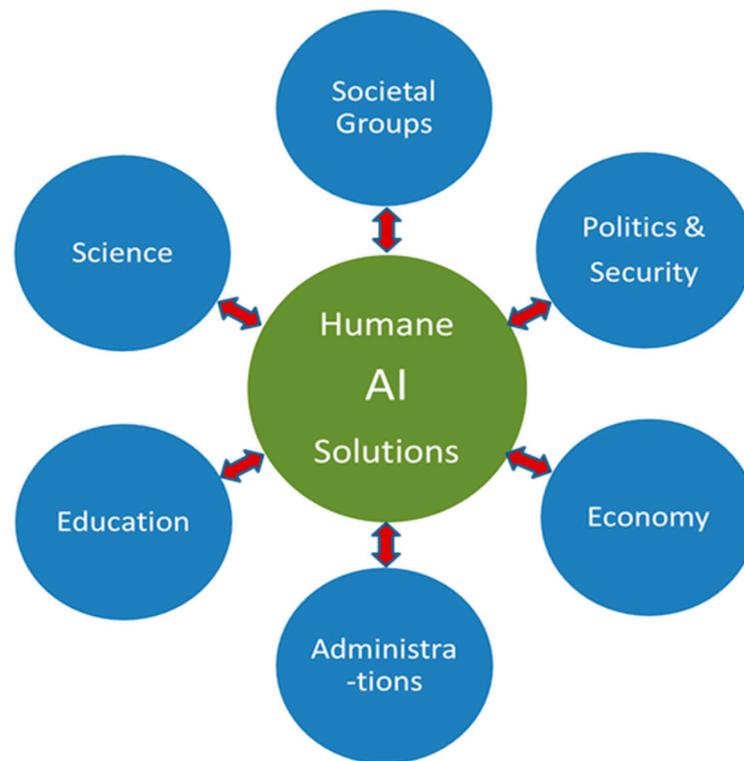


Figure 1. Cooperation model.

Developments may imply significant improvements to the human condition, but may also constitute risks of the contrary, including abuse or bias from groups with competing interests. Furthermore, the ultimate variants in future scenarios could include the risk of full takeover of societal governance by AI and robots, relegating mankind to a remnant of the past, enslaved or tolerated at best, or, at the other extreme, the creation of a full biblical Land of Plenty, with AI services and robots providing a paradise for humans to live in.

The novelty presented in this paper is that it takes a realistic view, while trying to develop an optimistic vision: a feasible strategy that would improve security, quality of life and human cohabitation, while maintaining dignity and self-determination and diminishing the fear and uncertainty of people, societies and politics when discussing how to manage the risks that may emerge from these concepts. The accompanying benefits could and should include climate control, optimal use of natural resources and the sustainability and recreation of biological diversity, along with improved overall safety and security and adequate risk and crisis management.

In a nutshell, AI is developing abilities that will allow for progress in many directions. The path towards which mankind strives varies substantially depending on region, cultural background, political system and principles, available resources, and environmental conditions. The manner in which politics, economy, science and societies prepare themselves for, agree upon and manage the elementary prerequisites for the transformation and shaping of future “cyber societies” will be essential. The basics need to be determined sooner rather than later, and this should be started now, at both national and global scales.

Section 2 provides a brief overview of how AI has developed over the decades, and shows that it is now at a stage where collaborative (CI) models between artificial and human intelligence are key to future societal applications and their expected effects, as discussed in Section 3. Section 4 discusses the impacts of these CI solutions on the economy, on politics, and on data requirements. Sections 5 and 6 discuss promising potential future applications of CI methods, while in Sections 7 and 8, the benefits and possible risks of such solutions are evaluated. As these ideas are rather challenging, the prerequisites are discussed and their feasibility is assessed in Sections 9 and 10, respectively.

2. Starting Point and Background

Having followed and being involved in public and scientific discussions and the professional development of AI for some 25 years, we believe that its potential has reached a level that will make a big difference in future applications. We, therefore, derive some ideas and draw some conclusions that could be useful for future debate and progress in the AI domain, and that hopefully will be substantial for future human society at large.

AI started some 60 years ago with the main intention to somehow understand human rationality and cognition technically, and to build human-level AI. In the 1980s, it developed into replicating human rationality and cognitive power in rather limited application areas. More recently, AI R&D has strived towards a solution to complex problems and generalizing theories and practices beyond the limits of human capability [1]. We thus put forward the question of whether AI and its extension to Artificial General Intelligence (AGI) is the intelligence of a machine that could successfully perform any intellectual task that a human being is capable of; see, e.g., https://en.wikipedia.org/wiki/Artificial_general_intelligence, accessed on 20 December 2020) should be pushed more, in the sense that it will design ways to solve problems in areas where human organizations are limited in speed, capacity, and ability, as well as manage the involved risks and create optimum solutions; in other words, problem areas which, in present political systems, are constrained through societal and political restrictions by the need for compromise and by the limitations in designing and executing sustainable long-term concepts. In this sense, after phases of both enthusiasm and nightmares (such as George Orwell: “1984”), we feel it is time to think of a positive far-end of AI.

Past public and political recognition and discussions of AI have occurred in waves. In the 1980s, hope and enthusiasm was concentrated on the prospects of extending computer ability to problems that are logically structured and rather deterministic (e.g., Expert Systems). The disappointments caused by the huge cost of the required computer power and limited scientific experience with complex logical networks and the self-learning ability of systems, etc., resulted in disillusionment. Only a few, rather modest, practicable solutions evolved and survived.

The present public discussion focuses more on what AI really means for our future society. Some effort has been spent defining what AI is really meant to be. Comparing AI with, and trying to simulate, Human Intelligence (HI) alone is, to some extent, frustrating, and may lead us in the wrong direction, at least as long as there are still so many loose ends regarding how human cognition and rationality really work, let alone their dependence on basic brain structure, genetically determined or not, for the interaction with sensory input, influence of mental/psychological states and the inter-operation of all these [2]. AI research and brain research are delivering a number of insights that will help to improve AI approaches and efficiency (e.g., neural networks, interfacing machines with biological neural systems, complex pattern recognition, language interpretation and many more). Scientific progress in artificial neural networks and machine learning have also contributed. We have a deep respect for the achievements reached with systems such as Deep Blue, AlphaGo/AlphaZero or Watson, and robotics that increasingly adapt or exceed human capabilities. However, our core question has arisen as to whether these should be the priority objectives of AI R&D and if they are the right approaches and exploitations for the more general benefit of mankind. Except for the sake of scientific progress, the core driver of and the related investments into RTD on AI should not primarily deal with, e.g., fancy dancing robots or playing games, to name only two of the most spectacular examples at present, which we, a bit disrespectfully, call “toys”.

We do not want to arrogantly disregard any achievements gained in AI to date. However, AI, in the future, from a political and socio-economic perspective, should focus more on solving the problems of mankind, societies, groups and individuals; in short, on improving the quality of life and common welfare. Compared to the “toys” referred to above, there are already fascinating and very helpful AI-based applications, such as pattern recognition (speech, face), language translation or improvements in self-driving cars. However,

what we are advocating here is to further focus investments, public attention and success stories on solutions that more strictly follow a humankind-oriented approach. AI should not concentrate too much on things that the human brain can do already, but on things that humans cannot perform, or perform too slowly or imperfectly, such as solutions to over-population, fairness in the distribution of resources and wealth, sustainable traffic and transportation concepts or genome sequencing for healthcare purposes. This, however, is a moving target, as ultimately AI may perform better than humans at almost everything. Nevertheless, there seem to be qualities in HI, such as “Emotional Intelligence” [3] or intuition, that may never, or only in an undefined future, be achieved by machine intelligence. Of course, this is not a black-and-white issue. Recognizing faces, automating production lines, improving security through tamper-proof identification, or optimizing the search for a parking lot have personal and economic benefits. The bottom line, however, of our admittedly rather fundamental considerations is that machine intelligence should no longer focus on imitating HI, but better capitalize on tasks that cannot, or can only rather sub-optimally, be handled by humans, in particular tasks for whole societies, politics, and large constituencies, such as the global financial system, commerce, trade, and alleviating boring work, risks, duties and suffering for mankind on a larger scale. “Too many problems (in most urgent real-world challenges) suffer from painfully slow progress” [4].

The potential of big payoffs for the advocated future AI applications should not be created via blunt automation alone, but involve cooperation between human and machine intelligence. It should not be a substitution for human intelligence and labour, but complement and improve humans’ mental and physical skills. In that sense, we will refer to CI. CI here stands for Complementary, Collaborative Intelligence, not Computer Intelligence as it is also sometimes used.

3. CI Applications Effects

Research and industry should concentrate CI on the exploitation of machine properties, such as basic computing power, fast networking and algorithmic concepts and solutions to complex problems. These CI application areas should be prioritized over those which obviously have not, or have very unsatisfactorily, been solved by humans to date, but for which promising solutions and payoffs are developing in AI (see also Table 1: Some Examples of Promising CI). From there, application domains should be identified where CI can produce help and solutions that will be significantly superior to human-made approaches tried in the past. The payoff will not just be high economically and commercially, but also in terms of enhanced societal qualities, including risk management.

Table 1. Some Examples of Promising CI (for further examples, see also [3]).

Subject Domain	Contribution by Human Intelligence (HI)	Contribution by Complementary Intelligence	Guardrails and Precautions Needed for Risk Management and Optimum Use
Medical diagnostics	Experience from dedicated cases; personal consultation	Systematic diagnosis based on large data bases; Improve focus and reliability of diagnoses	Privacy and data protection
Medical therapy	Control of therapy and results; Lessons learned feedback	Therapy suggestions from large data bases; robot support in, e.g., physical therapy; tele-therapy; reduce human effort in therapy (see, e.g., developing medical APPs)	Priorities and risk assessment; AI liability
Surgery and transplantation	Fallback and advice in critical situations; Cross-fertilization from other disciplines, e.g., materials research	Robot and tele-surgery support and even substituting human intervention; design and generation of synthetic body parts and organs; neuro-computer- interfacing	Insurance policy and accountability; help for the handicapped

Table 1. Cont.

Subject Domain	Contribution by Human Intelligence (HI)	Contribution by Complementary Intelligence	Guardrails and Precautions Needed for Risk Management and Optimum Use
Biotechnology	Conventional breeding; Concepts for accelerated evolution; survival strategies of biological life, diversity, etc.	Sequencing and optimization of genomes of food plants, animals, humans, and implementation control	Ethical criteria; rules on equality vs. preferences
Nanotechnology	Prioritizing and selection of candidate applications, e.g., in medicine; novel and “intelligent” materials design	Development of intelligent technologies, e.g., micro-robots and sensors, for continuous monitoring of physiological parameters and disease indicators; novel and “intelligent” bio-materials	Close cooperation between Bio/Nano and IT disciplines Methods for evaluating and supervising ethical principles
Social security and care systems, e.g., health insurance and pension systems	Transformation of political principles, social preferences, culture and constituents. Implementation of improved concepts and provisions for political acceptance	Development and control of sustainable mid- to long-term social security models. Incremental, periodic validation against reality	Confidentiality and data privacy; rules on fairness, equality
Climate Change	International agreement on facts, on objectives and action plans and their implementation	Suggested ways of optimum strategies and of compromising between diverging interests	Priorities; international frameworks and treaties
Ground and air traffic, Transportation, logistics	Implementation of improved concepts, related legislation and exception handling. New models of car (or other type of vehicle) sharing.	Concepts of information netting and automation (self-driving, alerting, parking management, airborne area logistics). Optimized mixes of public and private transportation and delivery systems. Automated modular, area-wide public transportation systems. Dispatching, logistical and financial optimization of resources	Rules for optimum and fair use
Education	Setting of new foci, e.g., on rationality, but also on societal, cultural and humane skills such as philosophy, arts, creativity and critical thinking and inquiring; supervision of data abuse, data quality control principles and organization. Avoidance of abuse of personal data and profiles and if fake information.	Provision, searching and distribution of contextual knowledge; Matching of personal profiles and academic and/or vocational education. Support of quality control of education results and of intelligent search, interpretation and validation of data and of using media. Software support of teaching and learning processes. Creation of symbiotic ideas for arts, e.g., fine arts, music.	Creating and supervising of new /changing education objectives and methods. Monitoring of societal impact.
Law and law enforcement	Regarding of individual circumstances in lawsuits, e.g., psychological factors, special societal situations. Basics on responsibility and liability of AI actors	Identification of and conclusions from applicable law. Identification of historical role cases. Care for equality of law enforcement/jurisdiction rules	Sharing of responsibilities between human and automata. Legislation regarding autonomous AI agents
Crime investigation	Handling of complexity issues and contradictions; credibility of witnesses	Tools for exploiting case based knowledge and forensics analysis; Improvement of credibility of facts and witnesses	Balancing between AI findings and human jurisdiction
Financial systems and taxation	Handling of special cases and exceptions. Enforcing of measures to early mitigate financial crises already at indication and warning status	Automation of taxation processes; Local and global crises to be indicated early and mitigation measures to be recommended. Models of financial behaviour in support of improved tax fairness and fraud avoidance	Fairness rules and governance legislation
Security and defence	Evaluation of basic political strategies and options. Interpretation of gaming results, setting of priorities and rules in decision making, procurement, etc.; principles of data security and privacy	Simulation of complex conflicts, wars, terroristic acts and other catastrophic scenarios (e.g., by gaming and exercising); analysis of results; robots in dangerous environments; Command and Control and decision support,	Trustworthiness, controllability, responsibility, accountability [5]; privacy and data protection, confidentiality rules

Table 1. Cont.

Subject Domain	Contribution by Human Intelligence (HI)	Contribution by Complementary Intelligence	Guardrails and Precautions Needed for Risk Management and Optimum Use
Socializing and partnering	Organizing and performing human exchange. Discussion of personnel selection parameters; psychological advises	Finding better matching partners; support groups of common interest. Intelligent partner matching of not only the “equal” but also the “different”	Data protection and privacy rules; psychological support
Job creation and assignment	Setting the basic personal preferences and assessing personal restrictions	Find the optimum job around the globe. Provide tradeoffs, e.g., between location, income, perspective, family situation.	Amelioration/harmonization of national, gender, racial, etc., restrictions and differences
Communications	Sharing of private and public authority over networks, and network services	Undecryptable coding, Real-time language translation, intelligent use of social networks	Control of information contents, e.g., of fake news, illegal contents, of profiling and abuse of algorithms
Global challenges	International agreements on objectives and control of efforts	Optimum strategies for climate control, exploitation of natural resources, bio diversity and sustainability; Suggestions for better governance of treating global challenges	Rules of fairness of distribution of benefits and of impacts
Managing the future	Identification of subject priorities, e.g., climate, bio-sphere, population, migration and involved problems; setting of priorities	Intelligent analysis and interpretation of mass data. Improved forecasting confidence and reliability; support in choosing priority subjects	Legal frameworks; international treaties and control mechanisms

Of course, related concepts (and, later, the solutions) need to be assessed against a variety of impacts they will undoubtedly have, including:

- E1 Compliance with Ethical standards;
- E2: Gross impact on Economy;
- L: Conformity with national and international legal settings and treaties, and possibly the need to modify them;
- P: Compliance with basic political agendas of “leading” nations and international regimes such as the UN, EU, OECD, IMF;
- S: The impact on and perception of such impacts by societies.

The basics of such an “EELPS” assessment have been developed in several international research projects and are further discussed in [6].

The superiority of CI then would mean the superiority of solutions, possibly superiority over human abilities, but also superiority over legacy organizations and political systems. Debating threat scenarios where AI substitutes the human (labor, cognition, rationality) “enjoys” huge attention from, and has a huge attraction to, the public and media at present (see, e.g., [7], strategy # 1; 7; 10). Instead of this, a positive and constructive concept of AI and CI should aim for solutions that are developed for, and implemented and operated in, cooperative models: here called Intelligent Complementary Human-Machine Collaboration (ICHMaC or CI). For the time being, we leave open the question of whether this will materialize in direct bio-machine interaction, organization of close cooperation, or the complementary use of brain and machine intelligence. The degree of integration of HI and AI will depend on the selected problem area and the related technical solutions being developed, as well on the risks and political will.

4. Validation and Consequences

4.1. Human vs. Machine Capabilities

When following this basic approach, there would be no need for further (mostly cramped) discussion on what is human and what is artificial. In CI concepts, they simply coexist and cooperate. This will, however, require a kind of general “societal contract” or agenda on the weighting of societal values and human skills relative to AI. As AI will become increasingly smarter, this is not a one-time fixed agreement; rather, it needs to

become a continuous process, including “sandboxes” for experimenting with all related societal groups, and clear criteria for the validation of benefits and risks.

There may be necessary limitations, beyond which AI has no say. Then, the somehow still fictitious discussion of whether AI is, or will ever become, superior to HI, and even endanger human existence, should become irrelevant. Inherent to the concept should be regulations to avoid false or counterproductive competition between AI and humans. False and counterproductive here means with new risks or insufficient compensation for those people or organizations affected (see also Sections 7 and 8 on benefits and risks). In the longer term, and from a more philosophical perspective, as in our concept, issues which are not so relevant today, but may possibly in the future, should be discussed, such as whether machines may ultimately acquire properties of consciousness, free will, a “spirit”, or whatever “mental” characteristics may be attributed to AI above or beside rationality. Attributing consciousness to AI systems, however, would require a clear definition and understanding of what consciousness really means, physically and logically [8]; this is not available at present, and it could take a long time to fully understand (see e.g., [9]). The consciousness of AI, then, would lead to extended elaborations on ethical impact of AI, and the evolution of superintelligence, transhumanism and digital “Singularity” [10], terms that form a different level of philosophy, which we will not further detail here.

However, these speculations will lose their menacing nature when societies realize the boons of AI and CI. “These challenges may seem visionary, but it seems predictable that we will encounter them; and they are not devoid of suggestions for present-day research directions” [11]. More on this topic is also available in [12,13].

Further practical pros and cons should be considered in socio-political discussions on future CI concepts, including:

- The goals of AI should not primarily strive to substitute human labour due to rationalization, but should rather liberate mankind from “stupid”, routine jobs, enabling humans to concentrate more on care and socializing;
- AI “machines” may act as autonomous, responsible and liable agents;
- AI actors may become morally superior to humans, e.g., in critical incident risk assessment;
- AI solutions may, and hopefully will, in many areas, interact with much lower failure rates than humans;
- AI should be free from negative emotional motivators which prevail in humans.

In a cooperative AI concept, as sketched here, there will be no need to fear risks of AI (CI) becoming adverse or hostile to humanity. The precautions of creating only “benign” systems and for preventing adverse, uncontrollable or dangerous AI solutions will be, and must be, an integral part of this CI, and of any AI concept [13].

CI will leave humans to complete those tasks that he/she can do better than a machine, which require abilities and skills which are hard to implement in computers (e.g., intuition, life experience, highly sophisticated skills and knowledge such as medicine, creativity, arts, and decision-making under contradicting arguments). This will be a dynamic process, as AI/AGI skills will develop and further capturing these human properties, and AI skill profiles and human societies’ trust in AI will evolve over time and integrate with human abilities to form advanced cooperative models, e.g., virtual reality arts.

4.2. CI and Economy

Economic and socio-political models of human–machine cooperation and partnerships, other than models of competition between human labour and machines, need to be created and established (sometimes called “Centaur Teams” with “Hybrid Intelligence”, e.g., in [3]), controlled and maintained. Interdisciplinary collaboration of the AI community with economics, and political, social, psychological and philosophical science will become indispensable in the design, development, implementation and operation of those models. Their benefits and shortcomings continuously need to be monitored and evaluated from these very different disciplinary perspectives (see also Sections 3 and 7).

A cost–benefit assessment (short-, medium- and long-term CBA) needs to be established and regularly applied. As these CI concepts will create, or at least facilitate, rather innovative business models and political scopes of action, classical CBA will no longer be adequate. New assessment methods need to be created that have to include in their equation—besides the monetary/materialistic payoffs—the socio-economic, socio-political and even ethical implications of CI concepts in society [6].

4.3. CI, Politics, Taxation

As we have seen in the industrial revolutions of past centuries, automation is key to progress. First, industrial machines replaced raw labour, then refined machines produced goods, then general-purpose industry robots could diversify and adapt quicker to market needs, then the computer helped further automation, then the advent of the internet (not an automation tool in itself, but a productivity booster) and, finally, AI, have allowed for the creation of an increasing number of goods and services, much more cheaply than was possible with human labour.

Any country not investing in AI technology or trying to slow it down will seriously fall behind, as did any country that missed the past industrial revolution, probably even more so. In the future, computers in general, and AI in particular, will dominate the economy of every developed country.

Some believe that AI will take our jobs and cause mass unemployment and poverty. Others believe that, for every job taken over by AI, new jobs will be created. We believe that both sides of the debate are misguided. The primary goal is not to have a job and for businesses to maximize profit, but to gain a decent standard of living for as many people as possible. Un(der)employment should not be the focal problem; one may even argue that it is desirable and affordable (for more rationale see e.g., [14]). Crucial to maintaining or increasing global wealth is that the total amount of quality goods and services that are created stays the same or increases, respectively. If a robot or an AI replaces a human worker, the same amount of goods and services are produced, and likely even more, or fewer, but with higher quality. It is then “just a political matter” to redistribute these goods and services fairly, ensuring that nobody is worse off—most likely, everyone will be better off and, in the long run, much better off. The importance of immaterial benefits will increase. Few people seem to appreciate this simple economic fact today.

This redistribution can be done in a number of different ways:

- One could continue to decrease working hours for all (e.g., rather than one person working 40 h and one person having no paid job, both could work 20 h with the same salary as before, then 10 h, 5 h, etc.);
- Retirement age could be reduced, instead of the current trend of increasing it;
- We could have universal basic income schemes, funded by higher tax rates, e.g., for those still employed, on productivity and on unproductive financial transactions;
- The government could guarantee every shares in companies for every citizen (maybe initially in companies they have been laid off from), so that they profit rather than suffer from the automation;
- Robots/AI could be taxed, though one needs to be careful with taxing productivity gains to make sure innovation is not stifled. Preferably, one would only tax the negative side effects of a particular automation, e.g., the need for re-education and temporary unemployment costs of laid-off workers.

Probably, a well-thought-through combination of these and other measures, suitably adjusted to the increasing automation over time, could avoid any larger outfall, as we have also largely managed technological advances in the past. Of course, the right government and governance is necessary for this transition to be possible and peaceful, but if this occurred, nearly everyone would profit from it (see more in Sections 9 and 10). Generally, policies should be set and managed in a way that the poor could become richer and others, such as investors and entrepreneurs, would profit, but not unduly, from this new societal paradigm.

The replacement or improvement of human work with IT has been ongoing for more than half a century, with benefits for all, causing temporary layoffs but generally producing new jobs and societal and economic benefits [15]. Tendencies, intentions and plans, however, to replace human jobs with AI, without adequate compensation and tradeoffs between benefits and losses, need to be avoided through political and economic precautions, and substituted with ICHMaC concepts. Supporting social programs and lifelong academic and/or vocational training will need to accompany most career profiles. The primary goal of human existence will no longer be to have a job, but to have a decent standard of living. The required human working skills will shift from routine to creativity, from standard work, replaceable by robots, to higher quality work that exceeds the abilities of AI, that will not be bound to a fixed number of working hours, but will require high flexibility from employees, innovation and the abandoning of standard procedures. CI will alleviate the compulsion for gainful occupation. Unemployment must no longer be considered a problem; it may even become desirable and affordable! Politicians, economists and entrepreneurs, however, will need to adapt to these changes in societal paradigms and agree on different taxation and social security models that will avoid “arms races” and other risks of conflict between human and complementary machine intelligence.

4.4. CI and Data

AI/CI concepts will rely on the availability of data to learn from and use in their applications, usually vast amounts of data. This requires the preparation and assessment of available or emerging databases and the creation of the necessary:

- Organizational and technical structures to maintain and communicate the relevant data;
- Political/legal prerequisites, including those on data propriety and privacy regulations, access rights and responsibilities;
- Technical solutions to harmonize data from various sources;
- Verification and validation of data;
- Intelligent compression of large amounts of data.

Data compression and information theory are not only important for saving storage space and channel capacity for transmissions, they also play a key role in modern approaches to artificial intelligence [16].

We will not be able to discuss AI using complex data sources within this article. Nevertheless, it is common understanding that all practical and useful AI applications begin with data. As an example of AI in manufacturing, see, e.g., <https://www.forbes.com/sites/willemsundbladeurope/2018/10/18/data-is-the-foundation-for-artificial-intelligence-and-machine-learning> (accessed on 20 May 2021).

A typical classification scheme of “big data”, as seen today in the era of internet with social media, with cloud services and many others, is depicted in Figure 2.

The top categories in this picture, Velocity, Volume and Variety of data, are only three of the most important technical classification criteria for mass data characterization.

Further data characteristics need to include:

- Data sources and capturing, for instance, from sensors (cameras, RFID, etc.), organizations, individuals, software systems;
- Processes of capturing, transferring, structuring and storing data;
- Processing, analyzing and validating contents;
- Maintenance and quality management, including authenticity, correctness, reliability and scalability;
- Security and data protection, which are one of the most sensitive precautions needed (see also Section 6.2 on security);
- Socio-political aspects of data use; see, e.g., Elizabeth Denham (2017), big data, artificial intelligence, machine learning and data protection, Version: 2.2 [18];
- Scalability and durability of databases.

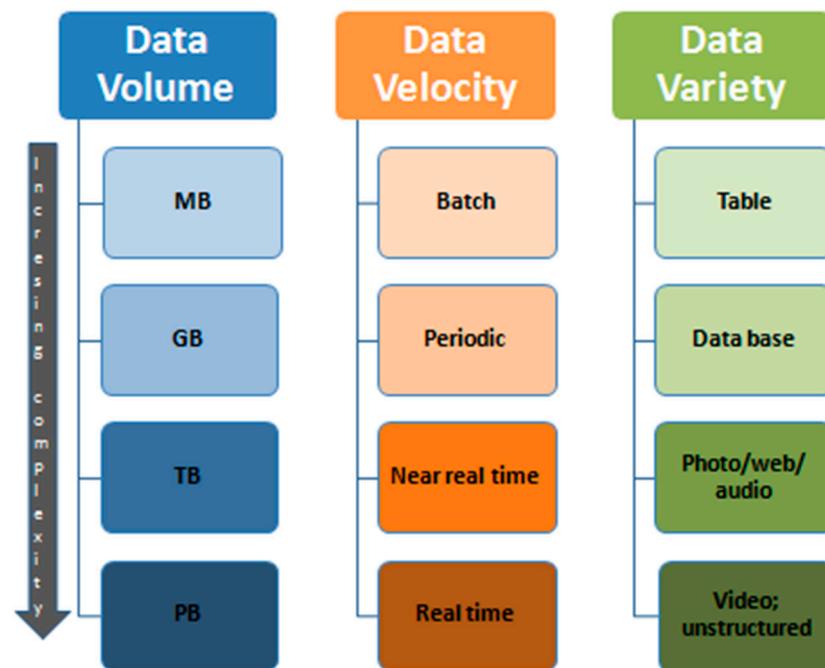


Figure 2. Big Data Classification Parameters (adapted from: [17] where, in addition to the Figure, a full suite of big data characteristics are defined).

Among this rich spectrum are the key requirements concerning data that will be used for AI. When we discuss the applications of AI/CI in the social context, as we do in this article, we believe that, for preparation at the socio-political level, certain elements are indispensable:

1. Public authorities: Public institutions and actors such as healthcare, security and transportation must agree and prepare to apply their data to novel solutions in AI;
2. Public–private cooperation: A consensus should be reached among public authorities with private big data resources. This, hopefully, will be reached in a consensus-creating environment. In the best case scenario, we hope for public–private win–win strategies, but reinforcement by regulations may become necessary in order to create and maintain a balance of power between governments and private industries;
3. International agreements, in particular when preparing for global AI applications, e.g., for security, climate control, population growth control, and the distribution of global resources or global rules for market and finance.

These efforts require political initiatives and need to be started now and sustained and supported by impartial research capacities.

5. Beyond CI

Futuristic visions are challenging and their risks need to be perceived and managed. This paper assumes the continued existence and (hopefully) the continued exploitation and improvement of political systems such as “western” democracies. Of course, changes may occur, even major ones, of constitutions, of priorities and rankings in economy, and of society and the political system, some of which will be recommended here. More revolutionary scenarios, however, such as singularities in the sense of [10,19] machine takeover or conscious machines [12,20], and their possibilities and probabilities, lie beyond the scope of this article.

Computers have always partially outperformed human performance by factors of millions and more, e.g., in calculation and communication speed, automation of office and production labour, and nobody has cared too much or would describe this as revolutionary rather than an evolution, mainly to the benefit of society (for many examples, see, e.g., [15]).

“AI Takeover” scenarios by “hyper-intelligence” and “superhumans” are likely to remain, for decades and maybe longer, in the category of Science Fiction. The basic question, at present but particularly in the future, is where to draw the line between technology’s blessings and curse, how to detect or forecast them and how to reflect on them. If the line drawn too narrow, we may forfeit chances for higher living standards—not just in terms of monetary gain, but also possibly a cure for cancer, the elimination of hereditary diseases, improvements in longevity, the prevention of war, and/or saving the environment and the climate. If drawn too wide, a “singularity” or other catastrophic developments may happen, the consequences of which we, and even their proponents, cannot assess and predict with the present way of thinking about the future.

6. Some Promising Examples of Future CI

6.1. Survey

The spectrum of applying AI in the CI sense, as discussed here, appears to be almost unlimited. The few selections of promising applications given in Table 1 are rather arbitrary candidates, however likely, and not exhaustive. Columns 2 and 3 approximately indicate how work could be shared between human and AI partners in different domains. This should demonstrate that the basic principle of CI is for HI and AI to complement rather than compete with each other. The chances, challenges and benefits are left to the judgement of the readers or experts in the subject domain. Requirements, risks and benefits in the implementation of CI will need detailed tradeoffs. However, some basic guiding rules for the implementation are mentioned in the right-hand column. Solutions and their realization will require extensive research and analysis, planning and consensus-building among all involved and/or affected stakeholder groups.

6.2. CI for Risk Management and Security

The present discussion about security comprises, in principle, three different aspects:

- (1) Security of AI/CI solutions themselves;
- (2) Risks from AI abuse or malfunction;
- (3) Security as a socio-political application domain for CI.

Here, we elaborate on the latter: security as a promising application area of CI, one of the many candidates shown in Table 1.

Security is the most fundamental need of humankind, following the basic physiological needs for food and water (<https://www.simplypsychology.org/maslow.html#gsc.tab=0>) (accessed on 20 May 2021); this is true for societies, groups and individuals. The protection and integrity of lives and possessions are deeply anchored in the psyche of mankind, due to evolution’s survival and selection principle, and are present in mankind’s organizational constructions. Treats are usually classified as natural disasters, technological risks and man-made disasters. The basic prerequisites for establishing and maintaining security are subsumed in the term Risk Management. A typical organization scheme of risk management is presented in Figure 3.

Almost all tasks in this activity cycle require considerable skills and resources. Despite various approaches to security and an enormous, historical trove of experience from wars and other disasters, precautions for and measures to maintain security have always been all but optimal. This is due to the huge complexity of the various problem areas, the fuzziness of data and the short reaction times required, as well as the diverging or conflicting interests of the individuals and groups involved, a lack of incentives for long-term planning in government and industry policies, and the tendency to ignore very rare but global, truly catastrophic events that could endanger humanity in its entirety [21].

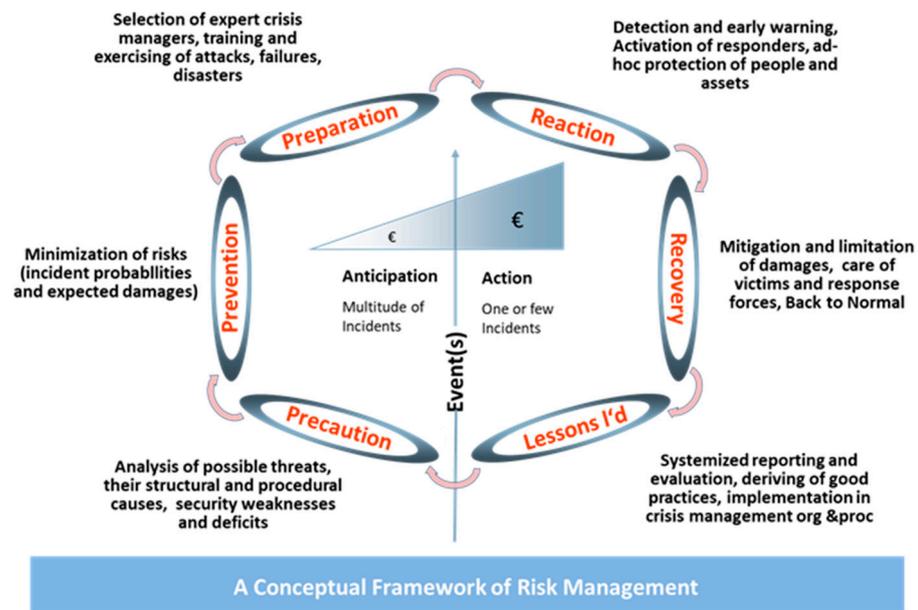


Figure 3. Security and Risk Management.

Even during the last 20 years, in which security increased in priority due to the 9/11 catastrophe and follow-up terrorist activities, measures taken to establish or modify organizations, in terms legislation and responsive actions, often appear ill-designed, or lacking in rationality. The authors appreciate all efforts taken and do not claim to know better, but with AI/CI, there is hope that we can do much better.

Nevertheless, the skills required in risk management, including data acquisition and analysis, organization, plans and actions, situation monitoring, and inferring and applying best practices, distinguish security as a preferred candidate for the extensive application of AI and CI. Arguments for this focus on security include:

- Security has to cover all areas of basic societal relevance, including ethical, social, legal, political and economic ones. Intelligent algorithms could improve data and privacy protection;
- Security requires a fast reaction in response to changing threats and disasters. AI will provide solutions to provide early warning, situation awareness and the fast creation of reaction options for responding units;
- These include scenarios of high complexity, with many uncertainties;
- Security will directly and visibly save lives and high-value assets, as well as improve well-being, both in reality and via fast-running, synthetic games and exercises;
- This will have a high positive impact on many other areas of society. CI security solutions will be appreciated and accepted in society, more than many other AI application areas;
- This has a high potential to increase trust and confidence in political systems and security measures;
- This will require intensive cooperation, trust and responsibility, from and between human and AI operations and resources (as has always been the case in security with technical resources such as rescue equipment or weapons);
- CI will substantially contribute to the reduction in risks for human reaction forces;
- The solutions will have to cover a huge variety of missions.

Overall, the delineated spectrum of tasks and potential make security a prime candidate for the exploitation of convincing AI/CI solutions. They could range from intelligent monitoring and detection and the assessment of warning indicators to the use of robotics instead of humans in dangerous areas; from optimal strategies to deal with pandemics to proven good practices in allocating patients to medical resources in disaster situations;

from intelligent analysis of operational situation data to optimal strategies for damage reduction and mitigation. There is an urgent need to proceed in this direction.

Elaborating on the examples of Table 1 would exceed the limits of this article and go beyond the focus of this special volume on Risk and Crisis Management. The scheme of CM discussed here, however, may serve as a role model for other subject domains listed in Table 1, many of which also address the security of societies in a wider sense.

As discussed at the very beginning of this chapter, security is a fundamental need. The other side of this coin, however, offers a huge potential for abuse: of all domains of political responsibility, and broadly independent of the political system in place, security implies the most traps and loopholes, to circumvent warp legal and societal control mechanisms. This is because the need for security often overrides the need for assessment by individuals as well as administrations, and limits rational discussions and the balancing of diverging interests and alternative options. Numerous such cases, particularly after 9/11, have occurred in security, leading to:

- Legislation conflicting with basic laws like constitutions;
- Unauthorized surveillance;
- Violation of data privacy;
- Breach of human rights;
- Abuse of data for unauthorized applications;
- Violation of basic ethical principles;
- Lacking proof of cost-efficiency and wasting money (for more examples, see, e.g., [22]).

Societies and politics need to support the preparation abilities of AI/CI in security to prevent rather than support the phenomena which are counterproductive to a prosperous and peaceful society, which is the aim of this contribution.

7. Evaluation of Benefits

Human brains and HI are wonders of evolution. Nevertheless, they also embody huge shortcomings that manifest in conflicting interests and controversies among institutions and decision makers, emotion-based preferences, and personal struggles with mental and/or physical force and injuries, which can lead to wars and global conflicts with apocalyptic outcomes, such as weapons of mass destruction, serious environmental impacts and social disruptions. The technical deficits in HI lie in the operating speed, replication accuracy of knowledge, diverging preferences of individuals and groups, limited objectivity, limitations in solving complex problems, and finding, introducing and sustaining optimal solutions to complex societal issues. Additionally, the limited sustainability and replication accuracy of human and societal memory and of the longer-term planning, control and maintenance of decisions and their impacts can cause problems. Religions (this in no way intends to diminish the beneficial effect of religion) and un-scientific or biased mindsets, conspiracy theories, etc., also contribute to the deficits in rationality [15]. Other shortcomings produce emotional drama or allow for the creation of ideologies that may be catastrophic for mankind or for groups and individuals.

Generally, the mankind's ability to anticipate, manage and mitigate risk seems to be inadequate, or at least suboptimal, in today's sophisticated, technological and organizational, highly developing world [21].

The CI approach presented here should concentrate on compensating for, or at least ameliorating or attenuating, these kinds of deficit. It is the basic postulation of this concept that this should produce the highest benefits for the well-being, standard of living, quality of life and the further development of societies into civilizations of the highest standards (some visions of what a "highly developed society/civilization" could actually mean are given in [23] (some 400 years ago) and [24]). Monetary profits are included, but should not dominate the proposed benefits. These should also entail the adequate control of climate, environmental exploitation and biodiversity. The possible benefits are obvious; however, they need to be verified and evaluated in a continuous process. Some promising research ends are given in the Stanford Human-Centred AI Initiative, stating that "the ultimate

purpose of AI should be to enhance our humanity, not diminishing or replacing it" [25]. Other initiatives follow similar principles, e.g., [14,26].

The benefits will be a mixture of material/monetary and immaterial/qualitative gains. Advanced methodologies need to be developed to assess the benefits, particularly life-cycle models and mixed qualitative/quantitative cost—benefit tools to balance the investments and benefits. This will also create new paradigms in the definition of “cost” and “benefits”. For example, if a “robot” can do what a human can for half the cost, it is “just” a political matter to ensure that everyone benefits, including those people that would otherwise lose their jobs and income.

In an earlier European Union project [27], the principles of a balanced assessment of measurable benefits, cost and societal impacts were developed. Different kinds of computer-based model focused on:

- Security risk assessment (probability of threats and damages);
- Life-cycle cost analysis (investment, operational, maintenance, update and replacement costs);
- Socio-political impact assessment, using mostly qualitative criteria (see Section 3).

These were defined, implemented, tested and successfully demonstrated in a number of use cases. The final challenge was to integrate the results on risk assessment, cost, and societal impact evaluation into an integrated decision support tool. Although, in that project, the applications were limited to investments in security, the basic methodology and plurality of the evaluation criteria can be viewed as a baseline, which can be modified to evaluate future CI concepts. Unfortunately, the willingness of public institutions to submit their decisions to such a systematic assessment was rather limited.

Nevertheless, using a comparable methodology, such wide balances between autonomy and controllability, as well as between investment, risks and benefits, need to be found for all CI solutions. This leads to the discussion of the risks of CI solutions.

8. Risks of CI

There is no technology without risk. The risks from fast-developing and internationally/global spreading technologies include those to and from AI, as AI will increasingly penetrate all domains of society worldwide. “Global problems need global answers” [28]. Apocalyptic nightmares about the future risks of AI are frequently publicized; however, they are usually biased and based on speculation rather than sound analysis and, therefore, are rarely helpful or likely to happen. Nevertheless, precautions need to be considered and implemented to ensure AI development and employment remains “benign” and avoid hostile or dangerous applications. An example of a potential risk is that deep learning AI system algorithms may one day have the potential to go beyond human control, or at least human comprehension of what they are actually doing, with the potential to abuse economic and political power, or even sabotage and undermine democratic constituents. Even worse could occur if such systems are hacked and used to cause military controversies or abused for large-scale criminal or terrorist purposes. This means that CI regimes will require internationally agreed-on standards, validation and certification rules, and a reliable regime of governance, before they are established. Some examples of CI applications and precautions are given in Table 1. Promising developing role models for governance can be seen in the European Union’s CIP, European Programme for Critical Infrastructure Protection policy (ECIP) (https://ec.europa.eu/home-affairs/e-library/glossary/european-programme-critical_en, accessed 20 May 2021) or in the General Data Protection Regulation (GDPR) of the European Union, which is compulsory for EU member states but appears to be becoming widely accepted as an international standard [29]. Another positive example comprises the agreements on cyber-security, which was first developed in the 1970s and has resulted in a number of internationally accepted standards in IT security (for the most prominent examples, see [30]). It should, however, be noted that the processes behind these initiatives took

decades, from the initial idea to putting the regulations into force; we may not have these timeframes with AI.

Political prerequisites, chances and obstacles regarding the further development of AI technology, and governance of CI solutions, are discussed in the next chapter.

9. Prerequisites for Success of CI Concepts

This CI concept could develop into a peaceful revolution of the human condition into the next generation of the socio-economy, related politics and modified constituents, and should be incorporated in state constitutions and national and international agendas (e.g., the UN). The basic vision of CI that should be created and communicated to all societal groups, organizations and their people is the creation of win–win situations between society and technology; between economic interests and those of the “working people”; between the maximization of consumption and the “higher values of life”; of a controllable financial sector with the aim to fairly balance the distribution of monetary resources among societies.

The classical neo-liberal concept (driven by capital, market and profit), to a large extent, needs to be replaced by, or transformed into, a concept of optimum benefit for societies (driven by social and humane values). The intent of this concept and its risks and problems is not, in any respect, to limit or hamper work on AI. Instead, it aims to employ and exploit AI to the mankind’s benefit. Defining what is “best” for humanity, however, is an even more complex, manifold, ambitious endeavor. Different cultures and political systems have different values and behavioral rules and processes; the interests of states or regions vary between nations, and often within nations. The same is true for different groups, be they the financial sector, associations of interest such as unions, religions, environmentalist movements and many more.

Nevertheless, we believe that the basic statements of the Universal Declaration of Human Rights of the United Nations (1948) and the European Convention on Human Rights (1953) could set basic guidelines, to be complemented by AI-specific aspects, e.g., those based on Isaac Asimov’s “Laws of Robotics” and adjacent principles, discussed by other great thinkers [31]. Without placing AI at the same level of risk as nuclear technology (civil and military), the international negotiations, treaties and organizations on nuclear control could provide useful lessons for a future international AI control regime. For a basic outline, see [21], chapters. 15 and 18. Beyond, and supporting, principles of humanity, cooperative CI models also need to define governance principles and roles for human–machine interfacing and partnerships, authority and hierarchy, management, and command and control loops.

Collaborative CI concepts will only prove successful if designed, developed and implemented in a joint effort of political, economic, commercial and societal forces, with massive support and participation from the scientific disciplines, including philosophy, sociology, psychology, economy, commerce and the Cyber and AI research community. This integrated concept needs to be established by an “elite movement”, furnished with resources and authority. Basic rules need to be developed and agreed on, starting now. It is, however, hard to imagine that the adequate organizations and processes could be implemented by present-day political, administrative and economic systems, and the people in power (a typical, negative example of such organization is the final disposal of high-intensity radioactive waste, and the GE example: political discussions and temporary experiments for 70 years, a promised time horizon of site selection in the 2030s, site establishment “starting in the 2050-ies”, i.e., roughly 100 years after the technology was introduced). A common understanding among all relevant groups will be necessary in all the discussed areas. Only a common agreement on the objectives of an ICHMaC strategy will facilitate solutions and better decisions than those of today. Political decision makers, legislation and responsible administrations need to accept the fact that CI can produce systems in support of many societal domains, superior to the possible solutions in present economic–political systems.

Consequently, the basic objective functions in the economy and politics need to be changed, to emphasize care for people and society, instead of saving traditional jobs, maximizing material wealth and becoming re-elected. Economic systems will need to focus on human-machine collaborative models, instead of maximizing profit at the expense of the “working”, as is common today in terms of a too-high divergence in income levels, opportunities and social states. The financial system needs to be regulated and controlled.

This all needs to begin with a modified vision of quality of life that coincides with the AI vision of relieving humans from the current primary need for jobs rather than to live much self-determined lives. This is not a “neo-communist” utopia in a new tapestry. Such basic ideas on quality of life were discussed by ancient Greek philosophers and in the vision of Thomas More [23] in the early 16th century. Present-day visions are also available, e.g., in [14]. This will, however, require new ways of thinking about societal constituents, modified or new models of creating and maintaining jobs and incomes, different concepts of education and lifelong learning that keeps pace with technological and societal developments, and modified or new concepts of general social security and care.

The income profiles of governments need to be adapted. Classical income taxes on wages and other earnings, consumption, etc., may shrink, and need to be ramped up in other ways, as discussed in Section 4.3. Alternative sources may be needed for unproductive speculation, real estate, the exploitation of rare natural resources, etc. Otherwise, the diverging distribution of income between the current, so-called “elite”, i.e., those who would “own” power over production, data, intelligence and capital goods, and the poorer and less privileged would accelerate, possibly to a ratio that may endanger social peace.

International organizations (such as the EU, UN, IMF, World Bank, OECD, G8 and global finance and commerce) need to support the ICHMaC concept. Political and scientific support will require increased funding of these organizations. In the same order of relevance, for the protection of the earth’s environment, biosphere, resources and climate, ICHMaC needs to establish an international planning regime and binding agreements, which will be sustainable over periods far exceeding the current electoral/legislative intervals and planning horizons.

10. Conclusions and Feasibility

The basic overall requirement must be insight into, and propagation of, the idea that CI will be able to overcome the flaws in past and present socio-political systems, and the consensus and will of “all” that better and collaborative solutions can be achieved and create unprecedented win-win-models.

Future solutions for society in the cyber world will require international, improved global rules and arrangements. This is true for the cyber domain including AI, as well as for climate protection and compliance with human rights. Unfortunately, there are always reverse tendencies, with populist movements prospering and propagating absolutistic, traditionalist illusions of “the past was better” agendas. With their limited appreciation of science and their backwards orientation, they will support the CI ideas. Instead, they will obstruct them.

Concerning the diligent treatment and exploitation of AI, we have only briefly outlined an option for the future. Its realization would need to start now, with an assessment of the prerequisites discussed above. Implementation of the concept or subsets of AI should be smooth and peaceful. However, realizing the pace of development in all cyber areas, including AI, we strongly suggest that politics and the economy prepare for such solutions sooner rather than later. Regarding the coming challenges and enormous potential of AI, present “western-type” governments are more or less dormant; real strategic thinking and political concepts are missing, and transferring progressive ideas into practice takes too long and is hampered by bureaucracies. “We are not a pre-emptive democracy, we are a reactive one” should be seen as a wakeup call [32]. The concepts discussed here are rather revolutionary. Therefore, policies need to follow precautionary and pre-emptive strategies. Unfortunately, more autocratic regimes, in principle, might be better-suited to enforce,

implement and sustain them, provided their basic objective and agenda is the creation of better conditions for all. However, historically, this has rarely been the case. Less desirable “motivators” for change could be global catastrophic risk (two examples: Fukushima; Coronavirus pandemic), forcing politicians to act. In any political framework, CI will require new and more efficient ways of consultation and consensus-building between peoples, governments, industry and research.

There are two extreme options, at the opposing ends of the spectrum, and many in-between. Things driven by AI, automation and robotics could cause deep tensions and further inequality in societies, and the government and economy will face increased pressure in a “3rd economic revolution”. Alternatively, politics and societies may become better-prepared for the substantial changes, and be ready to envision and plan a better world.

The first option may result in the further fragmentation of society, with people losing trust in existing political systems and technologies, feeling lost and no longer represented, needed, or taken care of. Increasing sympathies with the regional, autocratic systems and occurring at present are indications of this direction. At worst, this may lead to societal catastrophe, revolutionary mass movements, damage to democracy, and more (we have seen this in the 20th century and before).

The more optimistic second option, which is hopefully more likely and, of course, possible, would require unprecedented effort from political and economic sectors to create solutions guided by the prerequisites discussed above (and more). This would need to start now [33]. Unfortunately, there are almost no signs in politics, particularly not in education, that this will be prepared in time. As long as the high-level political discussion—as in Germany—is concentrated on internet access bandwidth, some local and rather modest investment in AI R&D, or the question of whether and how to allow or forbid smartphones in schools, we are tickling the little toe of the giant.

Speaking of giants, the largest ones at present, such as Google, Apple, Facebook, Amazon, Microsoft (GAFAM) and more, e.g., in China, are moving ahead. They have visions of future “cyber societies” and invest enormous amounts, including in AI (see e.g., [34]). Time will tell whether or not their results benefit mankind and individual societies. The truth is that their (GAFAM’s) power is increasing, while that of democratic governments is shrinking, although we are all aware that IT and AI are THE drivers of the present and future economy and everyday life. In any case, the future role of AI/CI in societies cannot be left to the forces of the market.

This may sound like pessimism, with two apocalyptic options. Nevertheless, through the influential progress of scientific and technological development, political systems will either be forced to adapt to new governance models and rules, or industry power will increasingly dominate, or even dwarf, the decision space of government plans and action. Seneca had a fitting recommendation for such a phenomenon 2000 years ago: “*Fata ducunt volentem, nolentem trahunt*” (Fate leads the willing and drags the unwilling). Only strong coalitions between the IT giants, governments and international organizations will be able to create future CI-based societies, as stressed in this article. As HI and AI should be viewed as two complementary abilities, the political systems need to view societal regimes such as quality of life and the economy as integrated parts of the same medal, rather than competing ones, as typically occurs at present.

An easier transformation into CI—and transformation is what we are discussing—should follow the “Think big, start small” principle: beginning an “Evolution” with rather small, non-spectacular, and politically and societal non-controversial sample projects (e.g., legalizing self-driving cars, improving healthcare systems, further automating routine bureaucracy, optimizing traffic flows and public transportation, fully automating taxation), learning from them and incrementally evolving politics and societies into CI domains of a larger size and greater impact (e.g., in law enforcement, legislation, rationalized political decisions guided by science, optimized and customized schooling, autonomous systems for safety and security, control of the finance system). In fact, cooperative IT-based models,

on a smaller scale, are already on their way, e.g., in software, for better administration, or robotics in production lines and supply chains (The Australian, 25 Jan 2019, pg. 18 on the automation of warehouse supply chains: “. . . more sites are combining robots and humans.”). The current and oncoming AI evolution, however, will not work at a biological pace, taking hundreds of years or more for small changes to evolve. We will have to cope with big changes within a few decades.

As realist-optimists, however, we are sure that the human race and society will survive for a long time, even without CI. Nevertheless, the question is HOW. That is what we are trying to advocate for, with some ideas for the betterment of society.

Author Contributions: R.H.: conceptualization of article, evaluation of mid- to long-term effects of AI on, and its potential for, future society; M.H.: AI methodologies and models, overall validation. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: No institutional/ethical review required; no studies involving humans or animals.

Informed Consent Statement: No studies involving humans or animals.

Data Availability Statement: No supporting data beyond those of the references were needed.

Conflicts of Interest: The authors declare no conflict of interest existing.

Abbreviations

AGI	Artificial General Intelligence	
AI	Artificial Intelligence	
Capability	Quality subssuming intelligence and required time and resources	
CBA	Cost-Benefit-Analysis or -Assessment	
CI	Collaborative or Complementary Intelligence	
CIP	Critical Infrastructure Protection	
CM	Crisis Management	
EELPS	Ethical, Economic, Legal, Political, Societal	Criteria and methodology for socio-political evaluation
EPCIP	European Programme for Critical Infrastructure Protection	
EU	European Union	
GAFAM	Google, Apple, Facebook, Amazon, Microsoft	Here used to represent powerful IT industries
GDPR	General Data Protection Regulation	
HI	Human Intelligence	
ICHMaC	Intelligent Complementary Human-Machine Collaboration	Processes implemented in CI
IMF	International Monetary Fund	
IT	Information Technology	
MI	Machine Intelligence	Here used synonymously with AI
OECD	Organization of Economic Cooperation and Development	
RM	Risk Management	
R&D	Research and Development	
RTD	Research and Technology Development	
UN	United Nations	

References

- Chomsky, N. Where Artificial Intelligence Went Wrong. 2012. Available online: <https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637> (accessed on 5 May 2021).
- Pinker, S. *The Blank Slate: The Modern Denial of Human Nature*; Penguin: Random House, UK, 2016.
- Bush, B. How Combined Human and Computer Intelligence Will Redefine Jobs. 2016. Available online: <https://techcrunch.com/2016/11/01/how-combined-human-and-computer-intelligence-will-redefine-jobs> (accessed on 5 May 2021).
- Deep Mind Mission. Available online: <https://deepmind.com/about> (accessed on 25 December 2020).
- Fraunhofer Group for Defense and Security. Rise of Artificial Intelligence in Military Weapon Systems. 2020. Available online: <http://publica.fraunhofer.de/starweb/servlet.starweb?path=epub.web&search=N-596967> (accessed on 15 May 2021).
- Hutter, R.; Blobner, C. Assessing the Socio-Political Impacts of Safety and Security Measures. In *Safety Risk Management: Integrating Economic and Safety Perspectives*; DeGruyter: Berlin, Germany; Boston, MA, USA, 2020; Chapter 13; ISBN 978-3-11-063378-8.
- Chomsky, N. 10 Strategies of Manipulation. 2013. Available online: <http://www.99posto.org/english/10-strategies-of-manipulation-by-the-media> (accessed on 15 September 2020).

8. Chalmers, D. Philosophy of Mind. In *The Conscious Mind, in Search for a Fundamental Theory*; Oxford Paperbacks: Oxford, UK, 1996; ISBN 8601419222391.
9. Aaronson, S. *Quantum Computing since Democritus*; Cambridge University Press: Cambridge, UK, 2019; pp. 34–42.
10. Kurzweil, R. *The Singularity Is Near*; Penguin Books: New York, NY, USA, 2014.
11. Bostrom, N.; Yudkowsky, E. *The Ethics of Artificial Intelligence*; Cambridge University Press: Cambridge, UK, 2014.
12. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014; ISBN 978-0-19-967811-2.
13. Yudkowsky, E. Artificial intelligence as a positive and negative factor in global risk. *Glob. Catastrophic Risks* **2008**, *1*, 184.
14. Welzer, H. *Alles Könnte Anders Sein: Eine Gesellschaftsutopie für Freie Menschen*; Futurzwei, Stiftung Zukunftsfähigkeit: Berlin, Germany, 2020; ISBN 9783103974010. Available online: <https://futzurzwei.org> (accessed on 15 August 2020).
15. Pinker, S. *Enlightenment Now*; Penguin Random House: New York, NY, USA, 2018.
16. Hutter, M. Human Knowledge Compression Prize. Open Ended. 2006. Available online: <http://prize.hutter1.net> (accessed on 10 September 2020).
17. Available online: https://en.wikipedia.org/wiki/Big_data# (accessed on 5 March 2021).
18. Denham, E. Big Data, Artificial Intelligence, Machine Learning and Data Protection. 2017. Available online: <https://rcn.fpf.org/2017/08/28/big-data-artificial-intelligence-machine-learning-and-data-protection/> (accessed on 15 January 2021).
19. Hutter, M. Philosophical Issues. In *Universal Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2005; Chapter 8.6; ISBN 978-3-540-26877-2.
20. Hutter, M. Conclusions. In *Universal Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2005; Chapter 8.7.
21. Bostrom, N.; Cirkovic, M.M. *Global Catastrophic Risks*; Oxford University Press: New York, NY, USA, 2008; Chapter 15 and 18.
22. Snowden, E. *Permanent Record*; Macmillan: London, UK, 2019; ISBN 978-1-5290-3565-0.
23. Thomas, M. In *Utopia*; Reclam: Stuttgart, Germany, 2012; ISBN 978-3-15-018875-0. Latin/German Version, Original Source 16th Century. Available online: https://www.buchkatalog.de/product/4099276460840799620/Buecher_Romane/Thomas-Morus/Utopia (accessed on 31 May 2021).
24. Lem, S. Star Diaries, the eighth voyage. In *Which Ijon Tichy Represents Earth to Petition for Its Admission to the Congress of United Planets*; Harcourt Inc.: Orlando, FA, USA, 1976.
25. Stanford University. The Stanford Human Centered AI Initiative. 2018. Available online: <https://hai.stanford.edu/welcome> (accessed on 10 December 2020).
26. HIIG: Humboldt Institut für Internet und Gesellschaft. Available online: <https://www.hiig.de> (accessed on 20 May 2021).
27. ValueSec. EU FP7 Project on “The Value of Security”, Deliverables D5.3 and D6.3. 2011. Available online: <https://cordis.europa.eu/project/id/261742> (accessed on 15 May 2021).
28. Harari, Y.N. *21 Lessons for the 21 Century*; Penguin Random House: UK, 2018.
29. Hearings of Mark Zuckerberg, CEO of Facebook, before the US Senate and Congress (10.04.2018) and the EU Parliament, 22 May 2018. Youtube. 2018. Available online: <https://fortune.com/2018/05/22/facebook-mark-zuckerberg-eu-parliament/> (accessed on 3 April 2021).
30. Security Cyber Security Standards. Available online: https://en.wikipedia.org/wiki/Cyber_security_standards (accessed on 20 May 2021).
31. Available online: https://en.wikipedia.org/wiki/Laws_of_robotics#EPSRC/_AHRC_principles_of_robotics (accessed on 20 May 2021).
32. Koppel, T. *Lights Out*; Penguin Random House: New York, NY, USA, 2015; Chapter 10; ISBN 978-0-553-41998-6.
33. Burchardt, A. Die Zeit #52/2020. Aljoscha Burchardt: “Steckt dSie KI in Den Sandkasten” (emgl.: “Put Them AI into the Sandbox”). Available online: https://www.genios.de/dosearch?dbShortcut=ZEIT&ZG_PORTAL=ADWPresse&gclid=EAlaIqobChMI3M6ru6b28AIVWpnVCh3jPgIVEAAYASAAEgLOS_D_BwE (accessed on 20 May 2020).
34. Ethics of Artificial Intelligence. Available online: https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence (accessed on 20 May 2021).