Artificial Intelligence—A New Knowledge and Decision-Making Paradigm?

Lei Huang (ID) **and Walter Peissl** (ID)

1 Introduction

Artificial intelligence (AI) has been considered as one of the critical technologies to address the future grand social challenges in a global context (Kaplan & Haenlein, 2020). Major economies are engaging to promote the R&D of AI through substantial policy efforts (Margetts & Dorobantu, 2019). Machine learning, neural networks, natural language processing (NLP), smart robots, knowledge graphs, and expert systems are among the key technical sub-systems that construct the current AI technological paradigm (Cresswell et al., 2020; Yablonsky, 2019). However, AI also raises concerns with regard to risks for society—from fundamental ethical considerations, through impacts on democracy, to the labor market. These risks and opportunities call for scientific policy advice on the basis of interdisciplinary technology assessment (TA) activities.

Contribution to: Technology Assessment in a Globalized World—Facing the Challenges of Transnational Technology Governance.

L. Huang (✉)
Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlstrasse 11, D-76133 Karlsruhe, Germany
e-mail: lei.huang@kit.edu

Chinese Academy of Science and Technology for Development, No. 8 Yuyuantan South Road, 100038 Beijing, China

W. Peissl
Institute of Technology Assessment, Austrian Academy of Sciences, Bäckerstraße 13, 1010 Wien, Österreich
e-mail: wpeissl@oeaw.ac.at

AI is believed to represent an interruptive potential regarding economies and societies. The application scenarios of AI are understood to cover many social domains (Di Vaio et al., 2020; Fosso Wamba et al., 2021; Perc et al., 2019; Popkova & Sergi, 2020), for instance, autonomous driving in transportation, robotic surgery in health care. The interaction between machines and humans may evolve to a new paradigm that human beings are transforming into data beings (Breazeal et al., 2016; Dautenhahn, 2007a, 2007b; Sheridan, 2016).

Meanwhile, some of research has claimed that AI should be considered as a general-purpose technology that generates grand implications in all sectors of the economy and society. For instance, deep learning technology will not only create market profits for online platforms, but also provides high-efficiency tools for social governance. In addition, with the accumulation of applications in specific economic and financial domains, AI technology presents the potential to change social structures (Klinger et al., 2018; Rasskazov, 2020). Hence, to achieve a better understanding of the potential impacts and necessary governance, we aim to identify the critical research topic of AI from a TA perspective.

In this chapter, we sketch a picture of global developments in AI and discuss potential fields of application as well as the demand for TA. In order to do so, in Sect. 2, we provide a short overview of historical developments of AI, and identify stakeholders involved in current AI R&D. In Sect. 3, we describe major activities in the international policy arena. Section 4 is devoted to TA activities with regard to AI in Europe and beyond. In Sect. 5, we conclude with the demand for future cooperation in global TA on the issues at stake.

2 The Identification of Stakeholders in a Data-Driven Artificial Intelligence Context

AI is a rapidly growing domain based on developments since the 1950s (Fosso Wamba et al., 2021). In 1950, Alan Turing offered the preliminary concept of "thinking machines" on the level of human beings (Turing, 2007). Based on the level of development, there are three types of AI: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence (ASI) (Lorica & Loukides, 2016). ANI is also termed "Weak AI", due to the limited information processing capability of specific tasks (Yadav et al., 2017). AGI or "Strong AI" refers to machines which can perform on the level of the human mind, performing a general level of intellectual tasks (Roitblat, 2020; Wang & Goertzel, 2012). ASI refers to the cognitive performance of machines that will surpass that of human beings (Narain et al., 2019). Actual AI systems refer to the "Weak AI" level so far.

Although some academic definitions have been made, there is no universal definition for AI (Helm et al., 2020; Legg & Hutter, 2007). Due to rapid development in this domain, the definition of AI is continuously changing. According to the EU AI watch report, *Defining Artificial Intelligence 2.0: toward an operational definition and*

taxonomy of artificial intelligence, the definition from the High-Level Expert Group on Artificial Intelligence (HLEG) has been considered as an operational definition of AI, as follows (European Commission. Joint Research 2020):

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that given a complex goal, act in the physical or digital dimension by perceiving their environment through **data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.** AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behavior by analyzing how the environment is affected by their previous actions.

According to this operational definition, AI is a certain kind of information system for processing data to achieve the given goal of making the machine think or act like human beings (Russell & Norvig, 2009). Hence, for information systems, a "stakeholder" means *"participants [being individuals, groups or organizations who take part in a system development process] … whose actions can influence or be influenced by the development and use of the system whether directly or indirectly."* (Pouloudi & Whitley, 1997).

The definition of AI has evolved since Turing's research on machine thinking. The core of the definition of AI is to make the machine think and act like a human being, as far as possible. The technological system of AI is a complex system. However, the basic logic of AI definitions is that machines need to learn from data, which is provided by humans via information exchange. Based on the current technological nature of AI systems, information exchange works in a digital format, which is called "Data" (Al-Jarrah et al., 2015). To achieve this goal, scientists have been developing relevant sub-technological systems that include hardware, algorithms, and human–machine communication interfaces (Shin et al., 2016). For instance, computational capability has increased 300,000 times from AlexNet (2012) to AlphaGo Zero (2017). More powerful computational capability means more efficient data processing for AI development (Al-Jarrah et al., 2015; Jordan & Mitchell, 2015).

Although the current AI ecosystem is still far from Strong AI, it has presented great potential for digital society and the economy (Mahadevan, 2018). Most of the current AI-relevant policies highlight the importance of data resources for AI development (as shown in Sect. 3, 'The international policy discussion'). Data-driven technologies and application scenarios are contributing to the training of algorithms, machine learning devices, and human-AI feedback systems. For instance, a large scale of graphic data has been applied to train the visual recognition algorithm (Lu et al., 2018). Data-driven technological progress presents one of the most critical differences between the AI technology from Turing's era and current AI (Chen et al., 2019). In addition, data-driven AI is not only a technological challenge but also a grand social challenge (Alyoshina, 2019). Most of the data resources which are used for AI development come from human societies.

Hence, the stakeholders of AI are those participants that can influence or be influenced by the development and use of AI systems, directly or indirectly. Furthermore, based on the HLEG operational definition of AI, data processing is the essential

characteristic for current AI systems. The stakeholders of current AI system development participate in data processing either directly or indirectly. However, according to the HLEG definition, the data processing capability must achieve the specific purpose that makes AI systems gain the function of decision-making by adapting their behavior to interact with the real world. Based on this specific purpose, the stakeholders of AI will also be involved in the intersystem behavioral interaction of AI systems. AI systems are evolutional systems based on data-driven information and communication technology (ICT) systems. All of the stakeholders in an AI system participate—on different levels and with different impact—in the construction of the ecosystem that includes technology innovation, application scenarios, and human-AI interaction interfaces (Dayton, 2020; Samoili et al., 2020).

AI is now generating value for different kinds of stakeholders involved in the technology systems development and use. A multi-stakeholder perspective on AI also considers the benefits, which are related to the value-creation process, for both industry and society (Güngör, 2020). Consequently, in line with the comprehensive review of the HLEG operational definition and the value-creation of AI, participating multi-stakeholders directly or indirectly process data from technology, industry and society, in order to develop the AI system.

The current developments of an AI definition show a significant feature of data-driven innovation. Hence, basic research and technology applications of AI strongly rely on the allocation of data resources from the R&D of technologies, economies, and societies. In addition, more stakeholders should be involved in AI development in terms of variant contribution of data resources and values represented by these data. We intend to illustrate the data-driven characteristics of AI development for further discussion in the TA domain.

AI systems and technology development

AI systems are complex systems that are built by subsystems of technology. For instance, machine learning, expert systems, smart robots, knowledge graphs, and natural language processing. At the current stage of development, those subsystems are designed and built on one of the most important principles: to process a large scale of data resource with effective and economical methods to achieve knowledge exchange between the AI system and the physical world (Duan et al., 2019; Eisenstein, 2019; Gu et al., 2018; Hassanien & Darwish, 2021; Kusiak, 2017).

AI system and industrial development

AI systems have been widely applied in the industrial domain. For example, in business intelligence analytics, autonomous driving, and intelligent manufacturing. Combined with specific algorithms, current data-driven business intelligence analytics gain more potential capability to process a large scale of data from multi-stakeholders from the market (Corea, 2019). Data-driven AI systems will offer a more stable autonomous driving system based on the R&D of semi-autonomous vehicles (Huang et al., 2019). For intelligent manufacturing, data-driven technologies have been considered as the fundamental layer of the entire system (Feng et al., 2020).

AI systems and societal development

AI systems are merging with societal systems (Garcia et al., 2020). Society offers a large scale of data resources for the development of AI systems (Rohlfing et al., 2020). At the same time, AI systems have the potential to change the entire social structure (Fosso Wamba et al., 2021). Decision systems based on AI have been applied in the domain of health and social care with a large scale of data-driven technology or application (Cresswell et al., 2020). AI has been widely used in the development of data-driven education since the millennium (Baldominos & Quintana, 2019; Guan et al., 2020). Data-driven algorithms have been adopted for research regarding social welfare improvement, e.g., improving refugee integration, incident management and pandemic control (Bansak et al., 2018; Elvas et al., 2020; Esposito et al., 2021). Furthermore, artificial intelligence assistive technologies have been widely applied in communication, politics and marketing (Margetts & Dorobantu, 2019; Van Esch et al., 2019). AI systems have also been implemented in many private and public security applications, such as biometrics, predictive policing and predictive recidivism algorithms in the judicial system. However, all of these bear a tremendous potential for surveillance and are prone to biases (Dressel & Farid, 2018; NIST, 2019; O'Neil, 2016).

3   The International Policy Discussion

Policy discussions are a crucial component of research related to TA. According to this perspective, we developed a comparative analysis of AI policies in global contexts. We operated a comparative policy content analysis of AI policies across two dimensions: a country comparison and an application area comparison.

   For this chapter, we analyzed national-level AI policy documents, strategies and plans. The policy analysis covered major economies and societies and include agriculture, taxation, transportation, education, and science etc. The current policy debate on AI has attracted widespread attention worldwide. Many international organizations and institutes have constructed databases of policy research. For example, the European Union Agency for Fundamental Rights (FRA) constructed the AI policy initiatives (2016–2020) database at European Union level for the research project "*AI, Big Data, and Fundamental Rights*".[1] However, in order to develop a more comprehensive and integrated comparative analysis of AI policies in a global context, we adopted the *OECD.AI Policy Observatory Database (OECD.AI)* as the main source using more detailed sample data. This database contains more than 600 policy entries from major sectors, and gives an overview from different perspectives and sources—from statistics and national strategies, to assessment reports on economic and societal implications.

---

[1] Source: https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights/ai-policy-initiatives, Access date: 30/11/2021.

We analyzed these sources with regard to specific domains of AI impact, and in relation to countries and geopolitically relevant regions. We adopted the policy document content analysis methodology to operate the comparison and conclusion of the current pilot policy practices. We also adopted *Pandas* as the data cleaning package in *Python* to remove duplicated policy data.

3.1 Different Discourses in Different Regions

According to the OECD.AI, more than 60 countries have already released a national-level initiative, policy or strategy of AI development. After removing duplicates in the OECD.AI Policy Document Dataset, there are 576 remaining documents. In terms of the yearly budget range, there are 14 documents (2%) marked as "More than 500 M (million in USD)", 13 documents (2%) marked as "100–500 M" and 154 documents (27%) marked as "Not applicable".

The USA, China, the EU,[2] and the UK are the AI-advanced majorities in terms of the amount of scientific publications on AI from 20 areas, including economy, digital economy, transport, health, industry and education. In the policymaking domain, the USA released 47 national-level documents of AI policy, initiative and strategy. The EU released 44 policy documents. The UK released 39 national-level documents.

USA

The USA has constructed a comprehensive policy framework of AI development. In 2016, the USA released three important policy reports, including *Preparing for the Future of Artificial Intelligence, National Artificial Intelligence R&D Strategic Plan,* and *Artificial Intelligence, Automation, and the Economy*[3] by the National Science and Technology Council (NSTC). The three reports clarified the purposes of and targets for USA AI development. In February 2019, the former US president Trump addressed the *Executive Order 13859: Maintaining American Leadership in Artificial Intelligence.*[4] This document emphasized the importance of USA leadership in the domain of AI, with applications in the areas of economy, society, security, and the military. The USA did not release any document that directly mention a yearly budget above 500 million (USD) held within the OECD.AI database. However, two documents from the USA have the yearly budget range of "100–500 M". One is *Joint Artificial Intelligence Center (JAIC)* released by the Department of Defence

---

[2] In OECD.AI database, the European Commission is responsible for the policy documents from the European Union. For instance, https://oecd.ai/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-27087, Access date: 25/05/2021.

[3] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf, Access date: 25/05/2021.

[4] https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/, Access date: 25/05/2021.

(DOD);[5] the main objective of JAIC is to set up a department to enhance the performance of AI R&D in the domain of security and military. The other document was released by the National Science Foundation (NSF), *NSF AI Research Institutes;*[6] it is a national AI research institutes program to promote longer-term R&D to maintain U.S. leadership in AI.

EU

According to the analysis of the policy documents in the OECD.AI database, EU AI policies have focused on the upgrade by AI applications in various areas which include industry, manufacturing, health, and energy. The EU has promoted the cooperation of R&D in the AI domain by releasing policies and member state-level plans, setting up research funding, and building up a laboratory. For instance, to emphasize willingness to cooperate on AI, the EU released the *Coordinated Plan on Artificial Intelligence*[7] in 2018. Furthermore, EU AI policy has underlined the research of ethics and humanity. In March 2020, the EU released a white paper, *On Artificial Intelligence—A European approach to excellence and trust,*[8] to declare the solution of information transparency in AI development, data security, privacy protection and the regulatory framework. In April 2021, the European Commission released a *Proposal for a Regulation laying down harmonized rules on artificial intelligence,*[9] which is the first legal framework for AI.[10] In addition, the EU has also emphasized specific areas to promote AI applications based on the niches with EU advantages. In 2017, the EU released *The Report of the High-Level Group on the Competitiveness and Sustainable Growth of the Automotive Industry in the European Union (GEAR 2030)*[11] to enhance AI applications in the automobile industry.

UK

Aiming to become one of the most important AI innovation centers, the UK has also released a series of national-level strategies and plans to promote AI development. In 2017, the UK published the *Industrial Strategy: Building a Britain fit for the Future (White Paper)*[12] to set out the government's plan to create an economy to promote AI application that fits the national industry strategy. In January 2018, the UK released

---

[5] https://dodcio.defense.gov/About-DoD-CIO/Organization/JAIC/, Access date: 25/05/2021.

[6] https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505686, Access date: 25/05/2021.

[7] https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence, Access date: 25/05/2021.

[8] https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, Access date: 25/05/2021.

[9] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence, Access date: 25/05/2021.

[10] See Footnote 9.

[11] https://ec.europa.eu/growth/content/high-level-group-gear-2030-report-on-automotive-competitiveness-and-sustainability_en, Access date: 25/05/2021.

[12] https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future, Access date: 25/05/2021.

UK Data Trusts Initiative[13] to offer the independent stewardship of data to secure AI development; the UK government also emphasized the relationship between AI and the digital economy in this document. Also in 2018, the UK set up an *Office for Artificial Intelligence (OAI)*[14] to enhance cooperation between different government departments, ministries, and multi-stakeholders in the domain of AI.

China

In the OECD.AI policy database, there are only eight documents regarding China's AI development. However, these policy documents cover the most important areas, including the national plan, industry guidelines, laboratory construction, and the education plan for current AI application and development in China. In 2017, The State Council for the People's Republic of China released the *National New Generation AI Plan (*新一代人工智能发展规划*)* [15] which has been considered as the cornerstone policy for China's future AI development and application. This national level plan involved a comprehensive policy framework for China's AI development initiatives and goals in specific domains that cover R&D, industrialization, talent development, education and skills acquisition, standard-setting and regulations, ethical norms, and security. In addition, China's Ministry of Science and Technology (MOST) released the *Governance Principles for New Generation AI- Developing Responsible AI (*新一代人工智能治理原则——发展负责任的人工智能*).*[16] This initiative highlights that China is emphasizing responsible AI development with the eight principles of harmony, friendliness, fairness, inclusiveness, respect for privacy, security and controllability, shared responsibility, open collaboration, and agile governance.

Other East Asian countries

Japan and South Korea have also presented an active policy attitude to AI development. In 2019, Japan released the national-level *AI Strategy (AI*戦略*)*[17] to clarify that the motivation of AI development should respond to the critical social challenges, including the aging society and sub-replacement fertility. In December 2019, South Korea released the *National Strategy for Artificial Intelligence (*인공지능 국가전략*)*[18] with the yearly budget range "More than 500 M". In this policy document, South Korea not only emphasized global digital competitiveness by 2030, but also declared to create 455 trillion Korean Won (approx. 405 billion US dollars) of

---

[13] https://datatrusts.uk/, Access date: 25/05/2021.

[14] https://www.gov.uk/government/organisations/office-for-artificial-intelligence, Access date: 25/05/2021.

[15] http://fi.china-embassy.org/eng/kxjs/P020171025789108009001.pdf, Access date: 25/05/2021.

[16] http://most.gov.cn/kjbgz/201906/t20190617_147107.htm, Access date: 25/05/2021.

[17] https://www8.cao.go.jp/cstp/ai/index.html, Access date: 25/05/2021.

[18] https://www.msit.go.kr/SYNAP/skin/doc.html?fn=b94d1781d5ef394ac6a63e274d3949be&rs=/SYNAP/sn3hcv/result/, Access date: 25/05/2021.

economic surplus and to enhance the living standard for the entire society through AI development.[19]

South America and Africa

Meanwhile, South America and Africa are still on the lower level of AI development in terms of the comprehensive comparison of scientific publications and policy performance. Six South American countries, Argentina, Brazil, Chile, Columbia, Peru, and Uruguay have already released national-level AI-related policies. In 2019, Brazil presented the *Brazilian Artificial Intelligence Strategy (Public Consultation) (Consulta Pública da Estratégia Brasileira de Inteligência Artificial).*[20] This strategy document declared AI development implications on the economy, ethics, development, education, and jobs. However, most of these South American policy documents are still directly related to the digital transformations and data governance.

In Africa, there were only four countries, Egypt, Kenya, Morocco, and South Africa, that released national-level AI-related policy, initiative or strategy documents. In those policy documents, similarly to South American countries, AI policy content was only mentioned in relation to digital development and data governance.

International policy conclusion

According to this comparative review of countries and regions, AI policy practices show the characteristics of diversity in a global context. Major economies have developed AI policies with a strong focus on their own economic, social, and technological development characteristics. For instance, policies from the USA presented the demand of maintaining leadership in every AI application area. AI polices from the EU placed more emphasis on governance of ethics and humanity. China's national AI plans strike a balance between promoting the development of specific application areas of AI and comprehensiveness. To promote economic and societal development is the most important policy target for all AI policy practices.

3.2 Areas of Application

According to the OECD.AI Policy Observatory Database (OCED.AI), 20 policy areas are related to AI application. Within these 20 policy areas, the economy and the digital economy are the most significant domains in terms of publications. The OECD.AI policy areas cover the major fields of economy and society, and include environment, health, agriculture, transportation, science and technology, and social and welfare issues.

---

[19] https://oecd.ai/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26497, Access date: 25/05/2021.

[20] https://www.oecd.ai/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26729, Access date: 25/05/2021.

Based on the statistical analysis of policy documents and scientific publications within OCED. AI, the AI & economy area is the most significant domain of AI policy. Most of the AI policy documents involve economic development and boost. In the area of AI & economy, policy research and documents concern AI application in new business models, data analysis, information systems and management performance. The application of AI has a potentially tremendous implication on the economy (Pratt, 2015). Data has been considered as a critical resource for the market, and AI technology will enhance the performance of data resources (Mirowski, 2007). However, there is the possibility that potential benefits in efficiency and profitability go hand-in-hand with a great deal of automatization, which may replace a large part of the workforce and may lead to unemployment, poverty, and fundamental structural changes.[21]

Hence, the digital economy is also one of the most active AI application domains. Further advancement of AI has augmented the digital economy with significant implications for the specific policy domains (Watanabe et al., 2018). The AI & digital economy area covers multiple applications that include data governance, digital security, privacy protection, and communication networks. Based on the sharing of data resources, AI development could create more application scenarios for the disruptive knowledge creation mode for the digital economy (Holford, 2019; OECD, 2020).

AI development is also generating more application scenarios within specific economic areas that include finance, industry, entrepreneurship, investment, and employment. In the area of finance and insurance, governments have already noticed that the application of AI enhances financial data processing efficiency in terms of the dynamic market status-quo and risk assessment (Bahrammirzaee, 2010; Heaton et al., 2017; Palmie et al., 2020; Sharma et al., 2020). In the area of industry and entrepreneurship, the application of AI is currently accelerating the process of digitalization with emerging business models (Garbuio & Lin, 2019). In addition, AI may substantially affect the investment environment by the application of the self-improving AI system (Hall, 2007; OECD, 2021). AI-driven industry and entrepreneurship will also significantly change the labor force structure and the employment environment. The automation systems of AI will replace more employed people (Zhou et al., 2020a, 2020b). Hence, the AI-driven employment environment may need massive evolution in the current education system with higher skills teaching and learning (Roll & Wylie, 2016).

Currently, the autonomous vehicle is one of the most important AI applications in the mobility area (Stead & Vaddadi, 2019). The autonomous vehicle has been considered as the biggest transition of mobility (Lăzăroiu et al., 2020). However, the autonomous driving system is by no means an independent technological system, it is rather a complex system that merges with the social and economic system (Kassens-Noor et al., 2020). For instance, besides the high efficiency sensor data transmission support system, the autonomous driving system relies on updated traffic regulations and new public infrastructure construction (Huang et al., 2019; Raiyn, 2018). Furthermore, the cost and potential environmental benefits will also be essential dynamics

---

[21] There is a broad discussion on this issue, for details see Sect. 4.2.

of the development of the autonomous vehicle (Noruzoliaee et al., 2018; Vosooghi et al., 2020).

In healthcare, AI application enhances the performance of diagnosing disease and health risk assessment (Waring et al., 2020; Zhou et al., 2020a, 2020b). AI systems can process a large scale of personal data to facilitate personalized healthcare and precision treatments and medicine (Ali et al., 2021), for instance, improving patient care with precision medicine and mobile health, high-performance management of health systems, better understanding public health, and facilitating health research and innovation. (Abdel-Basset et al., 2021). With the large scale of data resource integration from different areas, AI may generate more health care application scenarios (Dwivedi et al., 2021; Price & Cohen, 2019).

In terms of the comparative study, we found that current AI policies promote AI as a fundamental technological ecosystem in every application area. AI policies from different application areas all mentioned the importance of data resources for the development of AI technology. At the current stage, AI technological ecosystem and application scenarios are highly dependent on the integration and utilization of various data resources. The development of AI is showing strong characteristics of a data-driven innovation. This crucial dependency on data for training issues as well as for actual performance of its function leads directly to the discussion of the potential risks of widespread AI application. In the following Sect. 4, we will look more closely at AI-related TA activities in Europe and beyond, and finally will uncover the most striking issues discussed with regard to the societal implications of AI.

4 TA Activities in the Field so far and Options for the Future

4.1 Analysis of EPTA-Activities

The policy documents discussed above provide an overview of diverse national AI policies, mainly with regard to promoting AI as a key factor for future economic (and societal) development. This section analyses studies and findings from TA institutions around the world. The main focus here lies on the—often unintended—societal and ethical issues.

To date, most institutionalized TA activities are located within Europe and the USA. However, there is increasing interest around the world in TA and TA-like activities. EPTA is a network of organizations doing TA within or for their respective parliaments at a regional, national or European level. Since the U.S. Government Accountability Office (GAO) became an associate member in 2002, EPTA has extended beyond Europe and currently comprises 23 parliamentary TA (PTA) institutes. This process of global networking between PTA institutions now includes Chile, Japan, Mexico, and South Korea (Peissl & Grünwald, 2021). EPTA provides

a joint website[22] which is based on a database of TA projects, reports, policy briefs and news. We undertook a title and abstract search for the period 2008–2021 using the search strings "AI, machine learning, algorithm, robot, social media, and democracy". Based on the results we identified 173 entries; 73 projects, 39 reports and 61 policy briefs.[23] Cleaning the dataset by removing duplicates, including news and policy briefs for projects where a report was also available, we ended up with 80 entries for analysis.

The data show that PTA institutions have been aware of upcoming technology and related questions since 2008. The abstracts of these 80 entries were analyzed and clustered depending on the respective primary focus of the report or project. These PTA activities were mostly associated with AI (35) and robotics (13) in general. Together with digitalization (6), autonomous systems (5), algorithmic decision systems (7), and algorithms (3), this covers 69 of the 80 entries. The remainder address emerging technologies (3), labor (2), democracy, social media, surveillance, 5G, quantum technology, and financial technologies.

In a second step we found a set of keywords associated with areas of research. These show the characteristics of (P)TA studies: they are often interested in the general overview (25) of a development and try to figure out impact in several dimensions. The second domain of interest was the labour-market and implications for the workforce (12), third is the aspect of democracy (9), followed by ethics and healthcare (6 each). The legal framework for AI (4) and mobility and education (3 each) bring the total keyword analysis to 68 out of the 80 included entries. The other 12 deal with sustainability, the pandemic and space (2 each) and specific themes like surveillance, robot maintenance, quantum computing, precision farming, drug production, and consumer protection.

4.2   Main Areas of Discussion

In this section we will discuss some of the main areas of research on AI and society undertaken by PTA institutions. As already mentioned, there is a lively discussion about the effects of AI, automation, robots, and digitalization in general on the labor market. An overview of the different approaches in EPTA-member countries is given in the EPTA report of 2016 (EPTA, 2016). Unfortunately, a sharp distinction between the different domains is not possible, therefore the different areas are examined together. This is not problematic for our context, as AI can be seen as a basic technology for robots, automation, and digitalization of jobs beyond the production line.

The debate was triggered by a study by Frey and Osborne (2013), which showed that about 47% of all jobs in the USA were at high risk of being computerized within approximately 20 years. The main objection raised against this study's findings was

---

[22] www.eptanetwork.org.

[23] Search conducted 31.05.2021.

its historical analogy, which claims that all technological advances in history have also led to an increase in new (other) jobs. This argument, however, ignores the fact that economic conditions were different in earlier periods (Cas & Krieger-Lamina, 2020). With digitalization, we are facing completely new challenges, with ICT and AI flooding all areas of life, not only the production and service sectors.

The 2013 study has also been criticized for its methodological approach, but has been repeated in different contexts. An overview is given in Cas and Krieger-Lamina (2020), based on Lovergine and Pellero (2018). Although the high number of affected jobs was not replicated in other studies, the mere fact that the potential impact could be so large has led to discussions of the future labor market and unemployment. This attention may lead to less of an impact being seen in the future, as when we address the future, it is usually the beginning of shaping that very future.

The same applies to the area of social media platforms and their impact on societal communication, polarization and democracy. An EPTA-report provides an overview of different approaches to tackle the theme of "digital democracy" by EPTA-member States (EPTA, 2018). The digitalization of communication in the form of the internet originally gave rise to the hope that it would make democratization and broader participation possible. However, the developments of the last few years in the field of social media show that, in contrast, democratic structures and processes can be endangered. AI is not an insignificant contributor to this. The algorithms used by online platforms lead to a reinforcement of extreme positions and thus enable polarization in the spectrum of opinions, which makes constructive discourse more difficult. As shown by the examples of both campaigns in 2016 for the election of President Trump in the US and the Brexit vote in the UK micro-targeting on social media can be used to specifically approach voters and provide them with very different targeted content. In this way, general awareness-raising and fair information, by for example traditional media, are counteracted, and influence is exerted on these elections. Furthermore, there is a danger that these digital tools and infrastructure can be attacked and misused from the outside. This means that central elements of the political sovereignty of states can be undermined.

The relationship between new communication possibilities and AI-driven social media and democracy and the rule of law is of particular importance. TA-institutions at the European Parliament, in Germany, the Netherlands, Norway, Switzerland and the USA have conducted specific studies on this (Bieri et al., 2021; EPTA, 2018; GAO, 2020; Kind et al., 2017; Kolleck & Orwat, 2020; Marsden & Meyer, 2019; Neudert & Marchal, 2019; Tennøe & Barland, 2019; Van Est & Kool, 2017). Almost all other studies by EPTA members dealing with the effects of the widespread use of AI also address this fundamental challenge.

4.3   Responsibility, Transparency and Ethics

In more than a quarter of the analyzed TA-studies from EPTA members we can find basic issues like responsibility, transparency or ethics included in the abstract. These

issues are also discussed in the broader existing literature on AI and societal impact. It clearly shows that the most striking issue besides the direct effects on the labor market/workforce and communication and democracy are the more fundamental issues of responsibility, transparency and ethics. AI applications are intended to widely support decision-making or, even more riskily, autonomously decide upon certain actions. Processes triggered by AI systems may have impact on individuals or groups, their chances of societal participation, or even their very existence. So, it is fair to ask, who should be responsible for these decisions? In order to be able to locate problems or failures in algorithms, or other parts of the algorithmic decision-making systems (ADM), there is a need for transparency with regard to their internal mechanisms and the context of application.

Therefore, here we will present some lines of argumentation in depth. A prominent example of AI is its use in speech recognition and language processing. As the core element of digital assistants such as Smart Speakers like Alexa, Google Assistant or Siri, AI has found its way into many households and smartphones. Therefore, machines or systems have found their way into our households (and to some extent also to public offices) where they can not only comprehensively monitor our behaviour, but also influence it, from the way we communicate with machines, to how we communicate with each other, and how we move around our own homes. Since it is foreseeable that in the future, voice commands and thus digital speech recognition and language processing will be used in other areas (from cars, to systems' control in offices, and manufacturing), some fundamental questions arise: What is AI, what can it do and where are the limits to its application, if there should be any at all? What ethical principles guide considerations about what we want machines to do in the future, and what we don't?

For an ethical discussion, here we focus on an aspect of these definitions that comes from earlier approaches to AI. What does "artificial intelligence" mean? This is all the more difficult because there is no comprehensive, conclusive definition of "intelligence". For technical developments, the ISO established a definition for "artificial intelligence" in 2015. In this understanding, AI is the *"capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning."* (ISO/IEC JTC 1, 2015). However, "reasoning" and "learning" fall short of describing human intelligence comprehensively. In any case, AI research and development is trying to teach machines behavior that is modelled on, and as close as possible to, human behavior and decision-making processes. A distinction is usually made between weak and strong AI. In weak AI, algorithms solve individual, specific tasks, but do so quickly and, depending on the subject matter, to a very high quality. Examples include analyzing large amounts of data, pattern recognition, and predictions based on recognized patterns. Strong AI, on the other hand, describes a state where machines should have comparable intellectual skills to humans, and ultimately have consciousness similar to humans. However, this is primarily a visionary philosophical concept whose realization is widely doubted for the foreseeable future (Apt & Priesack, 2019). The discussions about a "superintelligence" which could ultimately prove superior to human intelligence and even

dominate humans, also fall into this area. Although this is a controversial topic, it is more likely to belong to the realm of science fiction.

From today's perspective, all available AI systems belong to the "weak AI" category. In addition to the above-mentioned advantages/abilities, they also have some fundamental deficits. These include a low capacity for abstraction, especially in the transfer of experience and learned knowledge to other contexts, high requirements for the pre-structuring of data, information and environments, and a lack of understanding and reasoning in the empathic sense. Even Alpha Go Zero, one of the most elaborate AI systems, is not able to do so. As a result, AI systems lack experience, tacit knowledge, judgment, empathy, and courtesy, as well as social learning and emotions that characterize humans and human intelligence. Or as Zweig (2019) puts it "an algorithm has no tact".

AI systems are designed to make more or less "autonomous" decisions based on available data and predetermined algorithms. When consequences for people or things result from these decisions, the question of responsibility arises. This usually becomes relevant in the case of negative consequences. So, who bears responsibility for the decisions made by AI systems? The systems themselves currently have no legal personality. However, responsibility is not only borne for negative damaging events. Responsibility is also derived to a certain degree from knowledge and competence. Therefore, one could also assume responsibility of such systems (which do not exist) for positive events. For example, what happens when a digital assistant hears someone calling for help, when children say they are being beaten, etc.? (Vlahos, 2019). Should a digital assistant then be required to call for help, or report assumed crimes? To this end, it is possible to discuss the question of when a certain type of reaction appears to be called for. This can be solved relatively easily by analogy with comparable situations involving humans. Appropriate behavior would then have to be programmed into the AI's algorithms. Much more fundamental, however, is the question of how a digital assistant comes to know about such situations in the first place. If it has been activated by one of the persons concerned, the system will listen in a permissible manner. However, since it could be argued that the digital assistants could be helpful in an emergency (sound the alarm), it would be conceivable to argue for their permanent activity. Here, however, an ethical conflict arises that points to the fundamental discussion of the security which can be brought by surveillance, and the cost of that security, in terms of, for example, loss of privacy. The decision is very much context-bound: under particularly threatening circumstances, such as in a hospital intensive care unit, we would like to be comprehensively monitored. But what about in everyday life? Do we want to be bugged everywhere so that we can potentially get assistance at some unknown point in an emergency (which will not necessarily occur)? How much "insecurity" can we humans stand, how much is reasonable for us, and how much security can actually be generated by additional surveillance?

AI systems are already integrated into many aspects of everyday life. They are interaction partners and also filters. Streaming services (Netflix, Spotify, etc.), Facebook, or even digital assistant providers such as Amazon and Google contain a large scale of personal data resources that can be applied to precise user portrait purposes

(Taffel, 2021); this means that they already know more about our personal preferences than our closest friends. Algorithms reinforce our purchasing decisions and solidify preferences. They thus determine the preferences of their users. In response to criticism of this, algorithms have been proposed which will make alternative suggestions in order to strengthen the sovereignty of users. But, this does not solve the problem, rather the opposite: they have an even more manipulative effect because they suggest an apparently objectified determination of preferences. But the result is the same in that the sovereignty of the user is—intentionally or not—drastically reduced (Stubbe et al., 2019).

Another problem in human interaction with AI (Amershi, 2019; Cai, 2019) arises from the attempts which are being made to make the systems as close as possible to the human way of communicating. However, no AI has yet passed the Turing test (Turing, 1950). Problematically, there is a strategic aspect of communication hidden in many applications of AI. That is, the systems try to feign a human counterpart in order to gain trust—which machines do not have a priori. However, this is not the same as two people talking to each other. Openness and transparency are important for the sovereignty of humans in interaction. Digital assistants have a similar effect, often docilely simulating an emotionless exchange with little controversy. Communication here loses its ambivalence, which humans know how to deal with through social experience, and through which we find out how we appear to others, and who we are to the other person. From an ethical perspective, this kind of communication with AI, which simulates a supposedly cooperative social interaction, whilst behind the scenes it pursues strategic purposes, is highly questionable (Stubbe et al., 2019). Current developments attempt to intentionally "enhance" the perfection of computers with human flaws and weaknesses. Unlike classic robot voices, Google Duplex, for example, inserts irregularities into sentences. Thus, apparent pauses for thought can be heard, or a "Mhmm" muttered now and then, together with abrupt pauses in speech. This gives the impression that the AI is responding to the conversation partner, or thinking (Kremp, 2018). Duplex was supposed to be integrated into Google Assistant on a test basis in 2018 (Herbig, 2018) and is now active in 49 states of the US and some other countries like Australia, Canada, India, Mexico, New Zealand and the UK (Callaham, 2021). Even if these applications currently only work in limited contexts and are trained for specific situations (hairdresser appointments, restaurant reservations), it nevertheless points to a fundamental ethical problem: users must be aware and able to know when communication with a machine is taking place.

A very different dimension of an AI application in communication is the MOODBOX smart speaker, introduced in 2016. It is primarily used to play music, but has built-in AI—called EMI—via the MOODBOX that is supposed to be able to detect the emotional status of the user from their utterances. The smart speaker checks how the owner is feeling and plays music to match the emotional state (Gineersnow, 2016). In this technology, of course, there is also great potential for the marketing of other goods and services, whose marketing/advertising is tuned to be easier to sell in certain psychological states.

With regard to emotion detection, we face some fundamental issues in AI and biometrics. The range of applications already include automated analyses of individual behavior patterns, such as a person's individual way of walking, as well as facial and emotion recognition, which are particularly controversial due to the enormous social risks. Emotion recognition is not only used in the advertising industry, but also to some extent in job interviews, in call centers and also in robotics/AI development (Masoner, 2020) (e.g. in the field of autonomous vehicles or for human–robot interactions). Toyota's new Concept-i series of automobiles is said to use AI systems and application of biometrics to recognize the driver's emotions by analyzing facial expressions and tone of voice. Thus, if the system (the A. I. Agent Yui) detects that the driver is stressed, it shall switch to autonomous driving mode ("Mobility Teammate Concept"). Alternatively, if the system registers signs of decreased alertness, the driver's sense of sight, touch and smell should be able to be stimulated to put the driver in a more alert state. In this way, the driver's stress level can be increased by smells, etc., but can also be reduced vice versa. The system is also designed to access a range of data from social media platforms, as well as activity and conversation content, in order to identify the user's preferences (Cheng, 2017). Furthermore, Toyota has announced that it has entered into a partnership with Microsoft, as have a number of other manufacturers in the automotive sector (Dudley, 2016). Emotion recognition is also controversial in AI research itself, both in terms of scientificity and meaningfulness, and a number of experts are calling for bans on emotion recognition (Honey & Stieler, 2020).

Machine ethics and ethics for AI

Machine ethics has been discussed frequently. It deals with rules in the field of AI and robotics. A distinction must be made here between ethical rules for AI systems and robots, and the so-called ethics of machines, or machine morality. The moral machine, which gives itself rules and then acts according to them, will only become relevant with the implementation of so-called strong AI, but is currently rather more the topic of science fiction than AI research.

The starting point for many considerations on the ethics of AI are the early robot laws of Isaac Asimov. These three laws, which Asimov developed in 1942, are supposed to counteract the robot's potential danger to humans by prohibiting actions (or omissions of actions) that cause harm to humans (first law), leaving the power of command over the robot with the human (second law), and ensuring the robot's self-preservation (third law). The laws are hierarchically structured, i.e. the third law may only be followed as long as the first two laws are not violated by it. However, as influential as Asimov's work was initially, it does not provide a sufficient and effective basis for the design of robots in general (Čas et al., 2017; Clarke, 1993, 1994).

Transparency as central requirement

The evocation of a potentially emerging "superintelligence" has certainly fueled the media hype surrounding ethical (and often dystopian) issues of AI. But there are also enough questions about the current state of the art and the expected further

development of weak AI that should be discussed and resolved in a broad societal discourse. This is all the more so as AI already plays an important role in people's everyday lives, and will play and influence many more in the near future. A central demand in this discourse is that of transparency, both about the use of AI (see above), and the modes of its operation, the built-in algorithms and their mode of action. Transparency is a necessary but not sufficient condition to control systems and thus build trust in them, which can also promote social acceptance. Society, and affected users or persons concerned, need to know how it works, but they also need a legal framework to be able to call for disclosure and sue for damages. There is also a need for institutions, which are able to both legally and technically deal with open questions from those affected. This leads to ongoing discourse on different levels of transparency and the respective means of governance.

Catalogue of ethical principles for AI

Besides transparency and (ex-post) evaluation of algorithmic systems there is a fundamental need to set standards for the whole development process from the very beginning. There are already a large number of initiatives of various kinds around the world that deal with ethical principles for AI. These include supranational associations such as the OECD, international professional associations such as the IEEE, and civil society initiatives.

A few examples illustrate the aims of these initiatives and their overlaps: fundamentally, the European Group on Ethics in Science and New Technologies of the European Commission published a declaration on artificial intelligence, robotics and "autonomous" systems in March 2018 (EuropeanCommission; DG Research & Innovation, 2018). This calls for the launch of a process that would pave the way for the development of a common, internationally recognized ethical and legal framework for the design, production, use, and governance of artificial intelligence, robotics, and "autonomous" systems. The declaration also proposes a set of ethical principles, based on the values enshrined in the EU Treaties and the Charter of Fundamental Rights of the European Union that can guide the development of this process:

- Human dignity
- Autonomy
- Equality and solidarity
- Responsibility, accountability
- Justice
- Democracy, rule of law
- Safety, security
- Physical and mental integrity
- Data protection and privacy
- Sustainability.

These fundamental principles should guide the development process on a meta-level, but there are also more specific requirements. In 2017 the Asilomar AI principles were developed—23 demands in the domains of research, ethics and values and long-term issues (Future of Life Institute, 2017). In 2018, the U.S.-based NGO

Public Voice published universal guidelines for AI (The Public Voice, 2018) and presented them at a major privacy conference in Brussels.

In 2019, a group of experts presented a policy paper with guidelines for the EU (AI HLEG, 2019). Based on fundamental rights and ethical principles, the guidelines list seven key requirements that AI systems should meet to be trustworthy:

- Human action and oversight
- Technical robustness and security
- Privacy and data management
- Transparency
- Diversity, non-discrimination and fairness
- Social and environmental well-being
- Accountability.

They are more or less equivalent to the dimensions raised in Public Voice guidelines. However, they go further by demanding the prohibition of secret profiling and making the general statement that governments must not generalize citizen assessment.

As can be seen from these lists, there seems to be a consensus that human dignity must be preserved, and humans should have ultimate control over systems. In turn, the resulting accountability can only be exercised if there is transparency regarding the modes of operation and algorithms. Transparency is also a basic condition for the establishment of efficient control systems, which are indispensable for effective regulation. In the context of global competition for the further development of AI, Europe tries to gain a quality advantage and thus a competitive edge with high ethical (and technical) standards. This is one of the main objectives of the proposed AI Act by the EU-Commission in April 2020.[24] Even though this proposed Act evolved from a consultation process, it does not cover all respective issues, deemed important from different stakeholders. However, it opens up public discourse on issues like high-risk AI systems, profiling, data protection, and biases in AI systems and ethics. This means that the limits society wants to set for advanced AI will have to be discussed. What do we want to delegate to machines, and what do we never want machines to decide?

According to Grimm (2018), "in order to use the Promethean potential of digitalization for a good life, we need … a digital value culture based on four pillars: (a) education and training (promoting ethical digital competence), (b) business and industry (value-conscious leadership, sustainable data economy), (c) research (interdisciplinary projects that bring together ethical and technological perspectives), and (d) political will (promoting value-based technology research)". In a global context TA could be a valuable partner in this enterprise, contributing on different levels.

---

[24] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

5   Conclusions on Practical Perspectives for Increased Global TA Co-operation and for Including Global TA in International Debate and Governance

This paper shows some of the hopes and fears regarding the widespread use of AI in more or less all areas of living. Visions of technological development, economic growth, higher efficiency and a better life are contrasted with a potential loss of workforce, poverty and dangerous development of cultures of discourse, which—combined with power inequalities—may lead to erosion of democracy.

AI as an emerging policy context has attracted widespread attention from governments, industry, culture and research in major economies around the world. Driven by policies in major economies such as China, the EU and the US, AI is gradually moving from basic research to more concrete application scenarios. However, there are still a number of issues and challenges before AI can become a truly general-purpose technology. Along its way, AI needs open discourse based on scientific facts, taking into account societal values. There is already input from TA institutions around the world on the specific features and potential impact of AI implementation. But there are open questions, which need translation from technological development to every-day life and back—and even more so on a global level.

As AI will influence nearly all areas of life, the context of use will be very diverse. In order to foster the development of responsible AI, **there is a need for a global ethical baseline for AI developers, implementers and users**. Elaborating general and globally accepted guidelines or codes of conduct need broad discourse and the participation of all those potentially involved together with affected stakeholders.

Transparency of AI systems is a fundamental prerequisite for accountability and control. There is a gap between understanding this requirement and the often mentioned "black-box" in AI, which supposedly makes it impossible to get full transparency over system-internal processes of decision-making. So, what kind of transparency is achievable, and what is needed? What does it mean if those demands don't match? Clarifying this in an interdisciplinary manner and communicating possible options to politics is a basic function of TA.

The diffusion and application of AI technologies has the potential to bring about tremendous changes to the social structure. The digitalization of human society will continue to be accelerated with the application and diffusion of AI technologies. However, AI as a new social resource will also be affected by the differences in resource allocation due to the previously existing problems in the traditional social structure. The development of AI will increase the potential risk of digital inequality on a global scale.

Responsible AI urgently requires global collaborative governance in a multicultural context. While AI has demonstrated the potential to become a general-purpose technology in a global context, the application scenarios of AI technology in specific countries and regions will be influenced by social, economic, cultural, and religious differences. Responsible AI should be based on the premise that it respects the diversity of people, and cultural and social settings, as well as the AI scenarios and

the diversity of governance concepts. Based on the diversity of governance concepts, all countries should enhance the transparency and credibility of AI development through collaborative governance in a global context. In addition, responsible data is the cornerstone of responsible AI.

The development and application of AI technologies will have a significant impact on geopolitical relations on a global scale. There are huge differences between developed and developing countries in terms of data resource allocation, basic research and development capabilities, and the level of industrial transformation. Accordingly, if AI is treated as a general-purpose technology, the realization of AI application scenarios will potentially create new global development imbalances that are detrimental to the achievement of the UN Sustainable Development Goals.[25]

Abundant empirical research is required to determine whether AI can be regarded as a general-purpose technology. As mentioned in the policy discussion in Sect. 3, current policies of application areas promote AI as a fundamental technological ecosystem in every domain. However, AI is still far from a general-purpose technology in terms of economics and societal performance. In addition, the development of AI strongly relies on the integration of related data-driven technologies. Data resource allocation is the key to determining AI as a general-purpose technology. However, a large amount of data resources (including personal data) is now controlled by giant online platform companies with significant market power. From a long-term perspective, data resource monopoly is not conducive to sustainable innovation in AI technologies. Therefore, both policy research and practice require an integrated view that data governance, platform governance and AI development be viewed as an intact ecosystem.

All of this calls for accompanying research and monitoring, as well as cross-cultural negotiation processes about desirable properties of AI systems and limits of application. TA can constructively contribute to this, if it succeeds in establishing corresponding global processes and institutions.

References

Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2021). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change, 163*, 14. https://doi.org/10.1016/j.techfore.2020.120431

AI HLEG. (2019). *Ethics guidlines for trustworthy AI*. Retrieved from https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. *Big Data Research*, *2*(3), 87–93. https://doi.org/10.1016/j.bdr.2015.04.001

Ali, F., El-Sappagh, S., Islam, S. M. R., Ali, A., Attique, M., Imran, M., & Kwak, K. S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems-the International Journal of Escience, 114*, 23–43. https://doi.org/10.1016/j.future.2020.07.047

---

[25] Source from: https://sdgs.un.org/goals, Access date: 14/01/2022.

Alyoshina, I. V. (2019). Artificial intelligence as a challenge for industries, economy and society. *TEDS'18*, 132.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., et al. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 chi conference on human factors in computing systems*, 1–13.

Apt, W., & Priesack, K. (2019). KI und Arbeit – Chance und Risiko zugleich. In K. Intelligenz (Ed.), *In* (Vol. Wittpahl, pp. 221–238). Springer.

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing & Applications, 19*(8), 1165–1195. https://doi.org/10.1007/s00521-010-0362-z

Baldominos, A., & Quintana, D. (2019). Data-driven interaction review of an ed-tech application. *Sensors, 19*(8), 1910.https://doi.org/10.3390/s19081910

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science, 359*(6373), 325–329. https://doi.org/10.1126/science.aao4408

Bieri, U., Weber, E., Nadja Braun Binder, Salerno, S., Keller, T., & Kälin, M. (2021). *Digitalisierung der Schweizer Demokratie – Technologische Revolution trifft auf traditionelles Meinungsbildungssystem*: vdf.

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). *Social robotics.* Springer-Verlag, Berlin.

Cai, C. J., Jongejan, J., & Holbrook, J. (2019). The effects of example-based explanations in a machine learning interface. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 258–262.

Callaham, J. (2021). What is Google Duplex and how do you use it? *Android Authority*. https://www.androidauthority.com/what-is-google-duplex-869476/

Cas, J., & Krieger-Lamina, J. (2020). KI und Arbeitswelt. In A. Christen, M. Mader, C. Cas, J. Abou-Chadi, T. Bernstein, A. BraunBinder, N. Dell'Aglio, D. Fábián, L. George, D. Gohdes (Ed.), *Wenn Algorithmen für und entscheiden: Chancen und Risiken der künstlichen Intelligenz* (Vol. 72/2020, pp. 144–164). vdf.

Čas, J., Rose, G., Schüttler, L. (2017). *Robotik in Österreich: Kurzbericht – Entwicklungsperspektiven und politische Herausforderungen* (ITA 2017-03). Retrieved from Wien: http://epub.oeaw.ac.at/ita/ita-projektberichte/2017-03.pdf

Cheng, M. (2017). Toyota's Concept-i Can Recognize Emotions Using AI and Biometrics. Retrieved from Future Car website: https://m.futurecar.com/1571/Toyotas-Concept-i-Can-Recognize-Emotions-Using-AI-and-Biometrics

Chen, L., Wang, P., Dong, H., Shi, F., Han, J., Guo, Y., . . . Wu, C. (2019). An artificial intelligence based datadriven approach for design ideation. *Journal of Visual Communication and Image Representation*, *61*, 10–22.

Clarke, R. (1993). Asimov's laws of robotics: Implications for information technology-Part I. *Computer, Band, 26*, 53–61.

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

Corea, F. (2019). *Applied artificial intelligence: Where AI can be used in business* (1st ed.). Springer International Publishing.

Cresswell, K., Callaghan, M., Khan, S., Sheikh, Z., Mozaffar, H., & Sheikh, A. (2020). Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: A systematic review. *Health Informatics Journal, 26*(3), 2138–2147. https://doi.org/10.1177/1460458219900452

Dautenhahn, K. (2007a). A paradigm shift in artificial intelligence: Why social intelligence matters in the design and development of robots with human-like intelligence. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), *50 years of artificial intelligence: Essays dedicated to the 50th anniversary of artificial intelligence* (Vol. 4850, pp. 288–302). Springer-Verlag, Berlin.

Dautenhahn, K. (2007b). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B-Biological Sciences, 362*(1480), 679–704. https://doi.org/10.1098/rstb.2006.2004

Dayton, L. (2020). Samsung's head researcher wants human-AI interactions to be a multisensory experience (Sebastian Seung talks about artificial intelligence). *Nature (London), 588*(7837), S129. https://doi.org/10.1038/d41586-020-03414-x

Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research, 121*, 283–314. https://doi.org/10.1016/j.jbusres.2020.08.019

Dressel, J., Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1), eaao5580. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5777393/. https://doi.org/10.1126/sciadv.aao5580

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data—Evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dudley, S. (2016). Microsoft to support new Toyota connected car company. *TheRecord*. https://www.technologyrecord.com/Article/microsoft-to-support-new-toyota-connected-car-company-54216

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., & Galanos, V. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 47.https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Eisenstein, J. (2019). *Introduction to natural language processing.* The MIT Press.

Elvas, L. B., Marreiros, C. F., Dinis, J. M., Pereira, M. C., Martins, A. L., & Ferreira, J. C. (2020). Data-driven approach for incident management in a smart city. *Applied Sciences, 10*(22), 8281. https://doi.org/10.3390/app10228281

EPTA (2016). *The future of labour in the digital era. Ubiquitous computing, virtual platforms, and real-time production.* Retrieved from Vienna: http://epub.oeaw.ac.at/ita/ita-projektberichte/EPTA-2016-Digital-Labour.pdf

EPTA. (2018). *Towards a digital democracy: Opportunities and challenges.*

Esposito, D., Dipierro, G., Sonnessa, A., Santoro, S., Pascazio, S., & Pluchinotta, I. (2021). Data-driven epidemic intelligence strategies based on digital proximity tracing technologies in the fight against covid-19 in cities. *Sustainability*, 13(2), 644. https://doi.org/10.3390/su13020644

EuropeanCommission; DG Research and Innovation. (2018). *Statement on artificial intelligence, robotics and "autonomous" systems European group on ethics in science and new technologies.*

Feng, Y., Zhao, Y., Zheng, H., Li, Z., & Tan, J. (2020). Data-driven product design toward intelligent manufacturing: A review. *International Journal of Advanced Robotic Systems, 17*(2), 172988142091125. https://doi.org/10.1177/1729881420911257

Fosso Wamba, S., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting & Social Change, 164*, 120482. https://doi.org/10.1016/j.techfore.2020.120482

Frey, C. B., & Osborne, M. A. (2013). *The future of employment: How susceptible are jobs to computerisation?* Retrieved from Oxford: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Future of Life Institute. (2017). *Asilomar AI principles.* https://futureoflife.org/ai-principles/

GAO. (2020). Deepfakes. https://www.gao.gov/assets/gao-20-379sp.pdf

Garbuio, M., & Lin, N. (2019). Artificial intelligence as a growth engine for health care startups: emerging business models. *California Management Review, 61*(2), 59–83. https://doi.org/10.1177/0008125618811931

Garcia, P., Darroch, F., West, L., & BrooksCleator, L. (2020). Ethical applications of big data-driven AI on social systems: Literature analysis and example deployment use case. *Information, 11*(5), 235. https://doi.org/10.3390/info11050235

Gineersnow. (2016). *This smart speaker knows how to feel your mood*. https://gineersnow.com/ind
ustries/audio-video/smart-speaker-knows-feel-mood

Grimm, P. (2018). *Grundlagen für eine digitale Wertekultur*. https://www.awo.org/unsere-arbeit/
theorie-und-praxis/tup-sonderband-2018

Gu, X., Yang, H., Tang, J., Zhang, J., Zhang, F., Liu, D., Hall, W., & Fu, X. (2018). Profiling web
users using big data. *Social Network Analysis and Mining, 8*(1), 24.https://doi.org/10.1007/s13
278-018-0495-0

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-
year data-driven historical analysis. *International Journal of Innovation Studies, 4*(4), 134–147.
https://doi.org/10.1016/j.ijis.2020.09.001

Güngör, H. (2020). Creating value with artificial intelligence: A multi-stakeholder perspective.
Journal of Creating Value, 6(1), 72–85. https://doi.org/10.1177/2394964320921071

Hall, J. S. (2007). Self-improving AI: An analysis. *Minds and Machines, 17*(3), CP6-259. https://
doi.org/10.1007/s11023-007-9065-3

Hassanien, A. E., & Darwish, A. (2021). *Machine learning and big data analytics paradigms:
Analysis, applications and challenges* (1st ed.). Springer International Publishing.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios.
Applied Stochastic Models in Business and Industry, 33(1), 3–12. https://doi.org/10.1002/asmb.
2209

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E.,
Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: Definitions,
applications, and future directions. *Current Reviews in Musculoskeletal Medicine, 13*(1), 69–
76.https://doi.org/10.1007/s12178-020-09600-8

Herbig, D. (2018). Google Duplex: Guten Tag, Sie sprechen mit einer KI. https://www.heise.de/
newsticker/meldung/Google-Duplex-Guten-Tag-Sie-sprechen-mit-einer-KI-4046987.html

Holford, W. D. (2019). The future of human creative knowledge work within the digital economy.
Futures, 105, 143–154. https://doi.org/10.1016/j.futures.2018.10.002

Honey, C., & Stieler, W. (2020). Expertenstreit über Emotionserkennung durch KI. https://www.
heise.de/newsticker/meldung/Expertenstreit-ueber-Emotionserkennung-durch-KI-4667496.
html

Huang, M., Gao, W., Wang, Y., & Jiang, Z.-P. (2019). Data-driven shared steering control of semi-
autonomous vehicles. *IEEE Transactions on Human-Machine Systems, 49*(4), 350–361. https://
doi.org/10.1109/THMS.2019.2900409

ISO/IEC JTC 1. (2015). *ISO 2382:2015 information technology—Vocabulary*. https://iso.org/obp/
ui/#iso:std:iso-iec:2382:ed-1:v1:en:term:2123770

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects.
Science, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities
of artificial intelligence. *Business Horizons, 63*(1), 37–50. https://doi.org/10.1016/j.bushor.2019.
09.003

Kassens-Noor, E., Dake, D., Decaminada, T., Kotval-K, Z., Qu, T., Wilson, M., & Pentland, B.
(2020). Sociomobility of the 21st century: Autonomous vehicles, planning, and the future city.
Transport Policy, 99, 329–335.

Kind, S., Jetzke, T., Weide, S., Ehrenberg-Silies, S., & Bovenschulte, M. (2017). *Social Bots*.
Retrieved from Berlin: https://www.tab-beim-bundestag.de/en/pdf/publications/tab-fokus/TAB-
Fokus-016.pdf

Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2018). *Deep learning, deep change? Mapping
the development of the artificial intelligence general purpose technology*.

Kolleck, A., & Orwat, C. (2020). *Mögliche Diskriminierung durch algorithmische Entschei-
dungssysteme und maschinelles Lernen – ein Überblick*. Retrieved from Berlin.

Kremp, M. (2018). Künstliche intelligenz: Google duplex ist gruselig gut. Spiegel Online. https://
www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-
a-1206938.html

Kusiak, A. (2017). Smart manufacturing must embrace big data. *Nature News, 544*(7648), 23. https://doi.org/10.1038/544023a

Lăzăroiu, G., Machová, V., & Kucera, J. (2020). Connected and autonomous vehicle mobility: Socially disruptive technologies, networked transport systems, and big data algorithmic analytics. *Contemporary Readings in Law and Social Justice, 12*(2), 61–69.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines, 17*(4), 391–444. https://doi.org/10.1007/s11023-007-9079-x

Lorica, B., & Loukides, M. (2016). *What is artificial intelligence?* (1st ed.). O'Reilly Media, Inc.

Lovergine, S., & Pellero, A. (2018). This time it might be different: Analysis of the impact of digitalization on the labour market. *European Scientific Journal, ESJ, 14*(36). https://doi.org/10.19044/esj.2018.v14n36p68

Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain intelligence: Go beyond artificial intelligence. *Mobile Networks and Applications*, *23*(2), 368–375.

Mahadevan, S. (2018). Imagination machines: A new challenge for artificial intelligence. *Paper presented at the proceedings of the AAAI conference on artificial intelligence.*

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature, 568*(7751), 163–165. https://doi.org/10.1038/d41586-019-01099-5

Marsden, C., & Meyer, T. (2019). *Regulating disinformation with artificial intelligence.* Retrieved from Brussels: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2019)624279

Masoner, A. (2020). *Wenn sich Maschinen für unsere Emotionen interessieren.* Retrieved from https://oe1.orf.at/artikel/673110/Wenn-sich-Maschinen-fuer-unsere-Emotionen-interessieren

Mirowski, P. (2007). Markets come to bits: Evolution, computation and markomata in economic science. *Journal of Economic Behavior & Organization, 63*(2), 209–242. https://doi.org/10.1016/j.jebo.2005.03.015

Narain, K., Swami, A., Srivastava, A., & Swami, S. (2019). Evolution and control of artificial super-intelligence (ASI): A management perspective. *Journal of Advances in Management Research, 16*(5), 698–714. https://doi.org/10.1108/JAMR-01-2019-0006

Neudert, L. M., & Marchal, N. (2019). *Polarisation and the use of technology in political campaigns and communication.* Retrieved from Brussels.

NIST. (2019). NIST study evaluates effects of race, age, sex on face recognition software. https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognitionsoftware.

Noruzoliaee, M., Zou, B., & Liu, Y. (2018). Roads in transition: Integrated modeling of a manufacturer-traveler-infrastructure system in a mixed autonomous/human driving environment. *Transportation Research Part C-Emerging Technologies, 90*, 307–333. https://doi.org/10.1016/j.trc.2018.03.014

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.

OECD. (2020). *OECD digital economy outlook 2020.*

OECD. (2021). *The digital transformation of SMEs.*

Palmie, M., Wincent, J., Parida, V., & Caglar, U. (2020). The evolution of the financial technology ecosystem: An introduction and agenda for future research on disruptive innovations in ecosystems. *Technological Forecasting and Social Change, 151*, 10. https://doi.org/10.1016/j.techfore.2019.119779

Peissl, W., & Grünwald, R. (2021). Parlamentarische TA. In S. Böschen, A. Grunwald, B.-J. Krings, & C. Rösch (Eds.), *Technikfolgenabschätzung – Handbuch für Wissenschaft und Praxis* (pp. 133–143). Nomos.

Perc, M., Ozer, M., & Janja, H. (2019). Social and juristic challenges of artificial intelligence (vol 5, 61, 2019). *Palgrave Communications, 5*, 1. https://doi.org/10.1057/s41599-019-0278-x

Popkova, E. G., & Sergi, B. S. (2020). Human capital and AI in industry 4.0. Convergence and divergence in social entrepreneurship in Russia. *Journal of Intellectual Capital, 21*(4), 565–581. https://doi.org/10.1108/jic-09-2019-0224

Pouloudi, A., & Whitley, E. A. (1997). Stakeholder identification in inter-organizational systems: Gaining insights for drug use management systems. *European Journal of Information Systems, 6*(1), 1–14. https://doi.org/10.1057/palgrave.ejis.3000252

Pratt, G. A. (2015). Is a Cambrian explosion coming for robotics? *Journal of Economic Perspectives, 29*(3), 51–60. https://doi.org/10.1257/jep.29.3.51

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

Raiyn, J. (2018). Data and cyber security in autonomous vehicle networks. *Transport and Telecommunication Journal, 19*(4), 325–334. https://doi.org/10.2478/ttj-2018-0027

Rasskazov, V. E. (2020). Financial and economic consequences of distribution of artificial intelligence as a general-purpose technology. *Finansy: teoriâ i praktika (Online), 24*(2), 120–132. https://doi.org/10.26794/2587-5671-2020-24-2-120-132

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., Esposito, E., Grimminger, A., Hammer, B., Häb-Umbach, R., & Horwath, I. (2020). Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems*, 1–1.https://doi.org/10.1109/TCDS.2020.3044366

Roitblat, H. L. (2020). *Algorithms are not enough: Creating general artificial intelligence*. The MIT Press.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall Press.

Samoili, S., Righi, R., Cardona, M., López Cobo, M., Vázquez-Prada Baillet, M., & De Prato, G. (2020). AI watch: TES analysis of AI Worldwide Ecosystem in 2009–2018, 30109.

Sharma, G. D., Erkut, B., Jain, M., Kaya, T., Mahendru, M., Srivastava, M., Uppal, R. S., & Singh, S. (2020). Sailing through the COVID-19 crisis by using AI for financial market predictions. *Mathematical Problems in Engineering, 2020*, 18.https://doi.org/10.1155/2020/1479507

Sheridan, T. B. (2016). Human-robot interaction: Status and challenges. *Human Factors, 58*(4), 525–532. https://doi.org/10.1177/0018720816644364

Shin, H. C., Roth, H. R., Gao, M. C., Lu, L., Xu, Z. Y., Nogues, I., . . . Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162

Stead, D., & Vaddadi, B. (2019). Automated vehicles and how they may affect urban form: A review of recent scenario studies. *Cities, 92*, 125–133. https://doi.org/10.1016/j.cities.2019.03.020

Stubbe, J., Wessels, J., & Zinke, G. (2019). Neue Intelligenz. In K. Intelligenz (Ed.), *neue Ethik? In* (Vol. Wittpahl, pp. 239–254). Springer.

Taffel, S. (2021). Google's lens: Computational photography and platform capitalism. *Media Culture & Society, 43*(2), 237–255. https://doi.org/10.1177/0163443720939449

Tennøe, T., & Barland, M. (2019). *Elections, technology and political influencing.* https://teknologiradet.no/wp-content/uploads/sites/105/2019/06/Elections-technology-and-political-influencing.pdf

The Public Voice. (2018). *Universal guidelines for artificial intelligence.* https://thepublicvoice.org/ai-universal-guidelines

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, Ausgabe, 236*, 433–460.

Turing, A. M. (2007). Computing machinery and intelligence (pp. 23–65). Springer.

Van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior, 90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Van Est, R., & Kool, L. (2017). *Human rights in the robot age : challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. Retrieved from The Hague:

European Commission. Joint Research, C. (2020). *AI watch: Defining artificial intelligence: Towards an operational definition and taxonomy of artificial intelligence*. Publications Office.

Vlahos, J. (2019). Smart talking: Are our devices threatening our privacy? *The Guardian*. https://www.theguardian.com/technology/2019/mar/26/smart-talking-are-our-devices-threatening-our-privacy

Vosooghi, R., Puchinger, J., Bischoff, J., Jankovic, M., & Vouillon, A. (2020). Shared autonomous electric vehicle service performance: Assessing the impact of charging infrastructure. *Transportation Research Part D-Transport and Environment, 81*, 15. https://doi.org/10.1016/j.trd.2020.102283

Wang, P., & Goertzel, B. (2012). *Theoretical foundations of artificial general intelligence* (1st ed.). Atlantis Press.

Waring, J., Lindvall, C., & Umeton, R. (2020). Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial Intelligence in Medicine, 104*, 12. https://doi.org/10.1016/j.artmed.2020.101822

Watanabe, C., Naveed, K., Tou, Y., & Neittaanmaki, P. (2018). Measuring GDP in the digital economy: Increasing dependence on uncaptured GDP. *Technological Forecasting and Social Change, 137*, 226–240. https://doi.org/10.1016/j.techfore.2018.07.053

Yablonsky, S. A. (2019). Multidimensional data-driven artificial intelligence innovation. *Technology Innovation Management Review, 9*(12), 16–28. https://doi.org/10.22215/timreview/1288

Yadav, A., Gupta, V., Sahu, H., & Shrimal, S. (2017). Artificial intelligence–New era. *International Journal of New Technology and Research, 3*(3), 30–33.

Zhou, G. S., Chu, G. S., Li, L. X., & Meng, L. S. (2020a). The effect of artificial intelligence on China's labor market. *China Economic Journal, 13*(1), 24–41. https://doi.org/10.1080/17538963.2019.1681201

Zhou, X. K., Liang, W., Wang, K. I. K., Wang, H., Yang, L. T., & Jin, Q. (2020b). Deep-learning-enhanced human activity recognition for internet of healthcare things. *IEEE Internet of Things Journal, 7*(7), 6429–6438. https://doi.org/10.1109/jiot.2020.2985082

Zweig, K. (2019). *Ein Algorithmus hat kein Taktgefühl – Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. Heyne.