Artificial Intelligence – Challenges and Chances for Europe

J O S E P H   S T R A U S

Max Planck Institute for Innovation and Competition, Marstallplatz 1, 80539
Munich, Germany. Email: j.straus@ip.mpg.de

*If we cease to believe in the future, the past would cease to be fully our past:
it would become the past of a dead civilisation.*

T.S. Eliot, *What is a Classic?* (1944)

As one of the building blocks of the fourth industrial revolution, artificial intelligence
has attracted much public attention and sparked protracted discussions about its
impact on future technological, economic and social developments. This contribu-
tion conveys insights into artificial intelligence's basic methods and tools, its main
achievements, its economic environment and the surrounding ethical and social
issues. Based on the announced and taken measures of the EU organs in the area
of artificial intelligence, the contribution analyses the position of Europe in the
global context.

Introduction

Together with analytics, cloud computing and the internet of things, artificial intelli-
gence (AI) forms an important part of the marriage of physical and advanced digital
technologies, which stands for what is commonly understood as the fourth industrial
revolution (European Patent Office 2017, 14). Industrial revolutions with their
manifold and unpredictable consequences, have always presented great challenges
to society. Because of the tendency of the rapid development of science and technol-
ogy, the very basis of such revolutions, to overwhelm and outdistance the law, legis-
lators are faced with the formidable problem of how to 'tame the unleashed genie of
science, so that it remains the servant not the master of mankind' (Markey 1989, 15).
However, AI technology has added a new quality to the old problem: it enables
machines, which use algorithms, to learn iteratively from data and think in concepts
and eventually turn themselves into a source of new knowledge, generated by AI.

Owing to its practically universal applicability, AI puts many long-standing paradigms into question and calls for new solutions. In order to let science and technology generate results that benefit the society at large, legislators more than ever have to interact with scientists, ethicists, economists and numerous stakeholders to reach responsible, prudent and farsighted future-oriented decisions.

AI – State of the Art

Since 1956, when the term AI was coined and defined as 'the science and engineering of making intelligent machines, especially intelligent programs' by John McCarthy, an American computer scientist (Gurkaynak *et al.* 2016, 753), AI has made remarkable progress. Today it is understood as 'a collection of technologies that combine data, algorithms and computing power' (European Commission 2020, 2), and in an exemplary manner it demonstrates its dependence on the interplay of techniques, discoveries and ideas. Computer scientists have developed a method of data analysis that automates analytical model building, which uses algorithms that iteratively learn from data, and which allows computers to find hidden insights without being explicitly programmed where to look, known as *machine learning* (ML) (Bratko 1994). Computer scientists have further developed the *deep learning method*, a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Thus, the computer gathers knowledge from experience. Therefore, there is no need for a human operator to interact with the computer (Kim 2016). The use of such methods in supercomputers such as IBM's Deep Blue with its enormous speed and storage capacity allowed the latter in 1997 to beat the chess world champion Garry Kasparov and signalled the advent of the first stage of AI development, namely that of 'Artificial Narrow Intelligence' (ANI). Although computer power is indispensable for the functioning of AI, even supercomputers such as Deep Blue, or the even much more powerful Chinese Tianhe-2 which can perform 34 quadrillion calculations per second and can solve complex problems extremely fast, have no perception of things other than the information provided to them by their creators.

Currently, AI is approaching the next stage of its development, namely that of 'Artificial General Intelligence' (AGI). AGI will represent 'Human-Level AIs', meaning that computers will be 'as smart as humans in every aspect and capable of performing all intellectual tasks humans can'. In its final stage, called 'Artificial Superintelligence' (ASI), AI will be 'much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills' (Gurkaynak *et al.* 2016, 751–753, referring also to Bostrom 2014). It would go beyond the skills of this writer to mention more than a few of the achievements that characterize the current status of AI. A very important opener of new AI applications is the use of 'Artificial Neural Networks' (ANN), an 'information processing paradigm that is inspired by the way biological systems, such as the brain, process information' (Stergiou and Siganos 2014). Shimon Ullman describes ANN 'as a highly reductionist

approach to model cortical circuitry' and observes that 'In its basic current form, known as "deep network" (or deep net) architecture, this brain-inspired model is built from successive layers of neuron-like elements, connected by adjustable weights, called "synapses" after their biological counterparts' (Ullmann 2019, 692). ANNs, like humans, learn by example. ANNs have

> the ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an 'expert' in the category of information it has been given to analyze. (Stergiou and Siganos 2014)

The enormous advantage of this approach is best demonstrated by the AlphaZero program, using a deep neural network, that recently convincingly defeated a world champion program in the games of chess and shogi (Japanese chess), as well as Go. The AlphaZero algorithm starts from random play and with no domain knowledge given, except the game rules. AlphaZero, unlike the state-of-the-art programs which are based on powerful engines that search many millions of positions, leveraging domain expertise and sophisticated domain adaptations, learns the necessary move probabilities and value estimates entirely from self-play and uses them to guide its search in future games (Silver *et al.* 2019).

The following selection of articles published in *Science*, *Nature* and *Cell* magazines may illustrate what AI technology has already achieved: 'A deep learning approach to antibiotic discovery' (Stokes *et al.* 2020); 'Computers turn neural signals into speech' (Servick 2019); 'AI mimics brain codes for navigation' (Savelli and Knierim 2018); 'FDA approves stroke-detecting AI software' (Report 2018); 'Machine learning classifies cancer' (Wong and Yip 2018); 'FDA backs clinician-free AI imaging diagnostic tools' (Ratner 2018); 'Computer-calculated compounds: researchers are developing artificial intelligence to discover drugs' (Fleming 2018); 'AI designs organic syntheses' (Lowe 2018); 'Machine learning for molecular and material science' (Butler *et al.* 2018); 'AI is selecting reviewers in China' (Cyranoski 2019); 'Artificial intelligence in cancer therapy' (Ho 2020); and 'AI shows promise for breast cancer screening' (Pisano 2020).

AI – Debated Aspects

AI, along with old-type concerns such as loss of jobs, safety endangered, discrimination generated, and the like, provokes also many new concerns such as loss of self-determination and self-control, or even the moral panic that AI would present a threat to society's capacity for empathy. Philosophers critically observed that, according to the logic of AI, there is no freedom of will, because machines follow a program. If they fail, this is due to the anomalies of the system. The same philosophers expressed concerns because some software-controlled systems have probabilistic functions, which assign a probability distribution of successor states to a state, rather than a fixed successor. As a consequence, 'learning' robots, meaning

probabilistic and not any more deterministic machines, can be constructed resulting in a suspension of the categorical difference between human and machine. 'The alternative is not a choice between determination and probabilism, but between determination and freedom' (Rümelin 2019, 12). Social scientists questioned the idea that AI's goal must be autonomy implying the conjecture 'that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it'. In their understanding, AI's goal and mission were not, 'and need not to be, to replace humans with machines, but to build machines that can work in tight interdependence with people.' This would require that 'engineers need to respect and learn from social scientists who are studying the complexities of human interaction with one another and with technology' (Cassell 2019, 30).

To what extent software engineers already cooperate with social scientists is difficult to assess; however, engineers are certainly aware that current ML algorithms, which are widely used in many sensitive areas, despite their potential to cause harm, still do not provide their users with an effective means for precluding undesirable behaviour. In order to provide relief for this serious problem, ML algorithms whose goal is not to produce a solution with a given set of properties, but to define the goal of the designer, which is to produce an algorithm with a given set of properties, were designed. Such ML algorithms, according to their designers, provide their users with the ability to easily (that is without requiring additional data analysis) place limits on the probability that the algorithm will produce any specified undesirable behaviour. They apparently do not just present a replacement for ML algorithms in existing applications but should also pave the way for new applications for which the use of ML was previously deemed to be too risky. Although these newly designed ML algorithms do not address the problem of imbuing intelligent machines with a notion of morality or human-like values, and do not solve the problem of avoiding undesirable behaviour that the user never considered (Thomas *et al.* 2019), they certainly constitute a great step forward in improving the safety of AI applications.

Because of its obvious interference with privacy, if applied in public areas, the use of AI in facial recognition technology from the outset has faced great scepticism and even decisive challenges. Since it has become known that the US company *Clearview AI, Inc.* has collected and stored in its private database three billion faces from predominantly publicly accessible sources, and developed an app whose computer code includes programming language to pair it with augmented reality glasses, the use of AI for facial recognition has provoked a strong negative reaction worldwide. The critics, among other things, pointed out that the algorithms 'were trained on and created by those with Caucasian-featured faces, which reinforces race-based biases in policing'. More than 600 law enforcement agencies in the US have already started to use the Clearview app. Among Clearview clients are also private companies. The US states of Oregon and New York, as well as a number of US cities, reacted to these developments and have banned facial recognition technology for policing and government use (Lipton 2020; DeVries *et al.* 2020). Now also the EU Commission is considering a temporary ban of 3–5 years on the use of facial recognition technology in public areas (Kelly 2020). In China, however, the use of AI for face recognition

also in public areas is omnipresent. The fact that one of the largest data factories in China, the private company known under the initials MBH, alone employs 300,000 workers across the country for labelling faces (the necessary preparatory work for face recognition), conveys some idea of the extent to which AI is used for face recognition in the country (*The Economist* 2020, 9).

AI – Economic Considerations

The economic importance for business and the actual use of AI technology transpire from a Special Report of *The Economist* entitled 'GrAIt expectation', published in 2018. Companies such as Johnson & Johnson, a consumer goods firm, and Accenture, a consultancy, use AI to sort through job applications and pick the best candidate. The casino and hotel group Caesars uses AI to guess customers' likely spending and on this basis offers personalized promotions to draw them in. The media and information firm Bloomberg uses AI to scan companies' earnings and automatically generate articles. AI makes it possible for the mobile operator Vodafone to predict problems with its network and with users' devices. Amazon uses AI for guiding robots in its warehouses and for optimizing packing and delivery, as well as for detecting counterfeit goods and powering its smart speaker/assistant, Alexa. Companies in every industry use AI to monitor cyber security and other risks. *The Economist*'s report emphasizes the 'Techtonic shifts' resulting from the use of AI. Without AI for product recommendations, targeted advertising and forecasting demand, tech leaders, such as Google and Amazon in the West and Alibaba and Baidu in China, would not be as big and successful as they are. Therefore, non-tech companies, worried that without appropriate use of AI they could be out-distanced, also started to acquire promising young tech firms. In 2017, firms worldwide spent about US$21.8 billion on mergers and acquisitions related to AI. Finally, the entire potential economic-value creation from AI in the next 20 years, according to estimates of McKinsey, will amount to some US$40 trillion (*The Economist* 2018, 3–5).

If one looks at patent statistics as an indicator for developments in AI-related technologies in geographic terms, then, as regards companies, among the top 20 applicants 12 are from Japan, three from the USA, two from China, two from Korea and one from Europe (Siemens, ranking 11th). However, IBM with 8290 and Microsoft with 5930 inventions are in leading positions. Of interest is also the fact that the Chinese Academy of Sciences (CAS) ranks 17th, and the Korean Electronics and Telecommunications Research Institute (ETRI) 20th (WIPO 2019, 60). Patent statistics for the top 500 applicants reveal high engagement of academic institutions in AI research and in patenting AI-related inventions. Among them are 167 universities and public research institutions, of which 110 are Chinese, 20 are from the USA, 19 from Korea, and four from Japan and Europe (of which highest ranking is the German Fraunhofer Institute at 159th, followed by the French Alternative Energy and Atomic Energy Commission in 185th position) (WIPO 2019, 16).

The dominant position of Chinese academic institutions, which not only represent some 60% of listed academic institutions but also occupy the first ten ranks, with CAS in the lead, does not come as a surprise. Peking University in Beijing first started to offer AI courses for undergraduates in 2004, and by now some 30 Chinese universities are heavily engaged in AI teaching (Cyranoski 2018). According to an analysis of the Allen Institute for Artificial Intelligence in Seattle, China's share of authorship of the 10% most cited AI papers has continuously increased and in 2018 reached 26.5%, behind only the United States at 29%, whose share is declining (O'Meara 2019).

China also has world-leading companies in computer vision, speech recognition and natural language processing, such as Magvii and SenseTime (privately owned start-ups for facial recognition, worth some US$4 billion and US$7 billion respectively – *The Economist* 2020), Unisound, iFlytek and Face++. The Chinese government is a big sponsor of AI technology. In 2018 it announced an investment of US$2.1 billion in an AI industrial park and it plans to become a world leader in the AI field by 2030 (O'Meara 2019). In response to Chinese AI developments, US Senator Charles Schumer, in a symposium held in Washington, DC on 5 November 2019, proposed that the US Government create an agency that over five years would invest US$100 billion on basic research in AI, 'To keep pace with China and Russia in a critical research arena and support work that US companies are unwilling to finance' (Merwis 2019).

AI – OECD Principles

In view of its economic, technological and social consequences, governments realized that the application of AI technology necessitates some internationally agreed ethical guidance. Therefore, on 22 May 2019, 36 countries, members of the Organization for Economic Cooperation and Development (OECD), and Argentina, Brazil, Colombia, Costa Rica, Peru and Romania, signed up to the following 'OECD Principles on Artificial Intelligence':

(1) AI should benefit people and the planet by driving inclusive growth, sustainable development and wellbeing;

(2) AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society;

(3) There should be transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes;

(4) AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed; and

(5) Organizations and individuals, developing, deploying or operating AI systems, should be accountable for their proper functioning in line with the above principles.

The OECD also recommended that governments facilitate public and private investment in R&D in order to spur innovation in trustworthy AI. They should foster accessible AI ecosystems with a digital infrastructure, technologies and mechanisms to share data and knowledge and create a policy environment that opens the way to deployment of trustworthy AI systems. People should be equipped with AI skills and workers should be supported to effect a fair transition and cooperation across borders and sectors to share information, develop standards and work towards responsible stewardship AI (OECD 2019).

In June 2019, China's National New Generation of Artificial Intelligence Governance Committee postulated harmony, fairness and justice, respect for privacy, safety, transparency, accountability and collaboration as ethical principles to control the area of AI development (O'Meara 2019).

Europe 'Discovers' AI

The European Commission also realized the necessity to adopt measures to cope adequately with the technological change generated by AI technology. In a Communication on 'Artificial Intelligence for Europe' on 25 April 2018[1] it announced it was devoting €1.5 billion to AI research funding through 2020. According to the Communication, for boosting the EU's technological and industrial capacity and AI uptake across its economy, the EU as a whole (public and private sectors combined) should aim to invest in AI research and development at least €20 billion by the end of 2020, and then €20 billion per year for the following decade. The Commission also offered a kind of 'inventory' of European AI-related achievements and capabilities and announced plans for a broad range of measures necessary to ensure that 'EU can make a difference – and be the champion of an approach that benefits people and society as a whole.' To achieve this goal, the Commission declared that it was time to make significant efforts to ensure that:

- *Europe is competitive in the AI Landscape*, with bold investments that match its economic weight. This is about supporting research and innovation to develop the next generation of AI technologies, and deployment to ensure that companies – in particular small and medium-sized enterprises making up 99% of business in the EU – are able to adopt AI.
- *No one is left behind the digital transformation.* AI is changing the nature of work: jobs will be created, others will disappear, most will be transformed. Modernization of education, at all levels, should be a priority for governments.
- *New technologies are based on values.* The General Data Protection Regulation was a major step for building trust, essential in the long term for both people and companies. This is where the *EU's sustainable*

---

1. Doc COM (2018) 237 final.

approach to technologies creates a competitive edge, by embracing change based on the Union's values enshrined in Article 2 of the Treaty on EU, i.e. respect of human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. As with any transformative technology, some AI applications may raise new ethical and legal questions, for example related to liability or potentially biased decision-making. The EU must therefore ensure that AI is developed and applied in an appropriate framework, which promotes innovation and respects the Union's values and fundamental rights as well as ethical principles such as accountability and transparency. The EU shall lead this debate on the global stage (EU Commission 2018a).

On 19 February 2020, the EU Commission published the 'White Paper on Artificial Intelligence – A European Approach to Excellence and Trust',[2] aimed at setting out policy options on how to achieve the twin objectives of promoting the uptake of AI and addressing the risks associated with certain uses of this new technology. According to the White Paper, Trustworthiness is a prerequisite for AI's uptake because AI, as a digital technology, is a central part of every aspect of people's lives; people should be able to trust it. The EU Commission emphasizes that a common European approach to AI is necessary to reach sufficient scale and avoid the fragmentation of the single market. To this aim, the White Paper contains two main 'building blocks': a policy framework setting out common EU measures necessary to mobilize public and private resources to achieve an 'ecosystem of excellence' along the entire value chain, starting in research and innovation, and to create the right incentives to accelerate the adoption of solutions based on AI, including by small and medium-sized enterprises; and a regulatory framework with key elements that will create a unique 'ecosystem of trust', ensuring compliance with EU rules, including the rules protecting fundamental rights and consumers' rights, in particular for AI systems operating in the EU that pose a high risk (EU Commission 2020, 1–3).

Europe in the Global AI Context

Prior to addressing the 'appropriate ethical and legal framework' envisaged by the Commission, a glance at the Commission's 2018 assessment of the *EU's Position in a Competitive International Landscape* seems necessary. Whereas the Commission admitted that Europe is behind in private investment in AI (€2.4–3.2 billion in 2016), compared with Asia (€6.5–9.7 billion) and North America (€12.1–18.6 billion), it claimed Europe is 'home to a world-leading AI research community, as well as innovative entrepreneurs and deep tech start-ups'. It emphasized that Europe accounts for the largest share of the top 100 AI research institutions

2. Doc COM (2020) 65 final.

worldwide, having 32 in the global top 100 as regards AI-related research paper citations (versus 30 from the US and 15 from China). It went on by pointing to Europe's strong industry, 'producing more than a quarter of the world's industrial and professional service robots (e.g. for precision farming, security, health, logistics)', and leading in manufacturing, healthcare, transport and space technologies, which all increasingly rely on AI (EU Commission 2018a, 4). While it is true that with KUKA, ABB and Comau, Europe has large and successful producers of industrial robots, that Siemens and Phillips are strong in healthcare, and Bosch in automotive applications, as evidenced in patent statistics (WIPO 2019, 60), European research institutions are, in practice, not among the applicants for AI-related patents.

Even more worrying than the poor patenting activity of European companies is that no single European company is among the world's leading digital companies – which have accumulated unprecedented cash piles (those of Apple, Google, Amazon, Microsoft and Facebook are equivalent to 10% of GDP in the US and 47% in Japan), market capitalization and volumes of information even on ordinary people, and increasingly control the infrastructure of the information economy (*The Economist* 2016). Even Microsoft has recently called on governments and companies around the world to share more data with other organizations to prevent what it warned would be a concentration of digital power in the hands of the US, China and a small number of giant tech companies (Waters 2020).

EU's Legal and Ethical AI Ramifications

The EU Commission in its 2018 Communication paid great attention to ensuring an appropriate AI legal and ethical framework. As regards the legal framework, the Communication pointed to provisions of the General Data Protection Regulation (GDPR)[3] already ensuring a high standard of personal data protection. Among others, it referred specifically to Article 22 (1) investing the 'data subject' with the right 'not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her' (EU Commission 2018a, 14). An accompanying 'Staff Working Document' on *Liability for Emerging Digital Technologies*[4] offered among other things an overview of safety rules relevant for emerging digital technologies at EU level, and addressed the principles of extra-contractual liability rules applicable in the same context. It further contained case studies dealing with AI power devices and systems (autonomous unmanned aircraft [drones] and autonomous cars), and the Internet of Things (smart home systems and cyberattacks), and pointed to aspects

---

3. Regulation (EU) 2016/679 of the European Parliament and of the Council of 17 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ EU No. L119/1, 4.5.2016.
4. Doc SWD (2018) 137 final.

of the 1985 Product Liability Directive[5] necessitating a further analysis. The annexed 'List of EU-legislation' reveals that the EU already has 64 directives and regulations dealing with liability, safety, etc.! The White Paper addresses new risks, which AI technologies present for users when the technologies are embedded in products and services, for instance as the result of flaws in the object recognition technology installed in an autonomous car, and which an improved legislative framework could address (EU Commission 2020, 12, 14).

The EU Commission paid great attention to also ensuring an appropriate ethical framework. Its statement that 'Spearheading the ethics agenda, while fostering innovation, has the potential to become a competitive advantage for European businesses on the global marketplace' (EU Commission 2018b, 17), makes it clear that 'ethics' is key to supporting 'secure and cutting-edge AI made in Europe.' Therefore, a High-Level Expert Group on Artificial Intelligence (AI HLEG), as an independent body composed of some 50 members from industry, universities, research institutions and business associations, was established. Its task was to draft: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations.

On 8 April 2019, the AI HLEG published 'Ethics Guidelines for Trustworthy AI'. The Guidelines are the sole responsibility of the HLEG and their use is voluntary. According to the Guidelines, a Trustworthy AI means an AI that should be *lawful*, complying with all applicable laws and regulations, *ethical*, ensuring adherence to ethical principles and values, and *robust*, both from a technical and social perspective, bearing in mind that even with good intentions AI systems can cause unintentional harm. However, the Guidelines do not deal with the aspect of lawfulness. To the Group's understanding, 'AI systems need to be *human-centric*, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom.' Therefore, the AI HLEG wants AI systems to get a competitive advantage by embedding Trustworthy AI in their products and services, which entails 'seeking to *maximize* the *benefits of AI systems* while at the same time *preventing and minimizing their risks*' (High-level Expert Group on Artificial Intelligence, AI HLEG 2019, emphasis in original).

Starting from legally enforceable fundamental rights such as: respect for human dignity, freedom of the individual, respect for democracy, justice and the rule of law, equality, non-discrimination and solidarity, and citizens' rights, the Group developed four ethical principles, called *ethical imperatives*, (1) respect for human autonomy, (2) prevention of harm, (3) fairness and (4) explicability, always to be adhered to by AI practitioners. For example, respect for 'human autonomy' requires that 'Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process' (AI HLEG 2019, 2–14). The Guidelines further provide a non-exhaustive list of requirements as guidance on the implementation and

---

5. Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, OJ EU No. L210/29, 7.8.1985.

realization of Trustworthy AI, which includes the following systemic, individual and societal aspects: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal wellbeing and accountability (AI HLEG 2019, 14–21). How AI can negatively affect diversity, non-discrimination and fairness, is explained by Yochai Benkler as follows:

> Algorithmic-decision systems touch every corner of our lives: medicinal treatment; mortgages and transportation; policing, bail, and parole; newsfeeds and political and commercial advertising. Because algorithms are trained on existing data that reflect social inequalities, they risk perpetuating systemic injustice unless people consciously design countervailing measures. For example, AI systems to predict recidivism might incorporate differential policing of black and white communities, or those to rate the likely success of job candidates might build on history of gender biased promotions. (Benkler 2019)

As regards 'human oversight', the Guidelines point to the possibility of achieving this 'through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach', i.e. through the capability of human intervention during the design cycle of the machine learning system and the human capability to oversee the overall activity of the AI system. 'Accountability/auditability', according to the Guidelines entails 'the enablement of the assessment of algorithms, data and design processes', which, however 'does not necessarily imply that information about business models and intellectual property related to the AI system must always be openly available.'

The Guidelines also provide for technical and non-technical methods for implementing Trustworthy AI. Whereas the non-technical methods – i.e. regulations, codes of conduct, standardization, certification, accountability via governance frameworks, education and awareness to foster an ethical mind-set, stakeholder participation and social dialogue and diversity and inclusive design teams – are known and already well-tested means for handling societally sensitive technologies, the envisaged technical methods are themselves an AI product. They should 'translate' requirements for Trustworthy AI 'into procedures and/or constraints on procedures, which should be anchored in the AI system's architecture'. According to the Guidelines, this could be accomplished

> through a set of 'white list' rules (behaviors or states) that the system should always follow, 'black list' restrictions on behaviors or states that the system should never transgress, and mixtures of those or more provable guarantees regarding the system's behavior.

Methods to ensure values-by-design would provide precise and explicit links between the abstract principles that the system is required to respect and the specific implementation decisions.

> The idea that compliance with norms can be implemented into the design of the AI system is key to this method. Companies are responsible for identifying the impact of their AI systems from the very start, as well as the norms their AI system ought to comply with to avert negative impacts.

Finally, the Guidelines provide a non-exhaustive Trustworthy AI Assessment list (pilot version) to *operationalize Trustworthy AI.* The list not only fills six pages with boxes full of questions, but also allocates responsibilities to a company's Management and Board, Compliance/Legal department/Corporate responsibility department, Product and Service Development or equivalent, Quality Assurance, Human Resources, Procurement and Day-to-Day Operations (AI HLEG 2019, 16-31).

It is too early to assess the impact these Guidelines may have on future development and implementation of AI in the EU and beyond. They certainly do not make the impression of being the creation of an 'industry-dominated "ethics washing"', as was apparently commented by an academic member of the HLEG (Benkler 2019). Rather the opposite: apart from the fact that the Guidelines are a voluntary set of rules, as for the substance, they enjoin strict ethical constraints on industry.

AI and Intellectual Property

In view of the huge public and private investments in developing AI-related technologies and the heavy patenting activities of all major players in AI, be it industry or academic research institutions or universities, reported above, the total absence of intellectual property aspects in the EU Commission's AI-related documents gives reason for serious concerns. Despite persistent doubts as to their economic benefits, intellectual property rights, especially patents, play a decisive role as incentives and discrete backing of long-lasting research efforts and are a very important means for securing high-risk investments (Straus 2017, 577–582). Eventually, patents provide for competitive advantages and secure markets. It is obvious that AI-related patents are sought after and, if granted, will fulfil exactly those aims. In other words, as the WIPO statistics show, US, Chinese, Japanese and Korean patent owners, who also control an enormous amount of data, the lifeblood of AI, will dominate the future European market of AI-related technologies, should European companies, the EU Commission and EU Member States not revise their attitude towards the use of patents in this area. At minimum, the EU Commission should soon address these issues as a high priority and link funding, wherever appropriate, to AI-related research results eligible for patent protection. It is worth noting that almost one third of US-patented inventions in 2017 relied on federal research investment (Fleming *et al.* 2019).

In the practice of the European Patent Office (EPO), AI-related inventions fall in the category of computer-implemented inventions and are patent eligible whenever the software makes a technical contribution, e.g. in the fields of medical devices, the automotive sector, industrial control, communication/media technology, automated natural language translation, voice recognition and video compression (EPO 2018a). However, when it comes to the special circumstances of how AI-related inventions are generated, e.g. by algorithms, which learn iteratively from data and experience and provide by themselves technical solutions, a number of issues questioning existing paradigms are still open and will have to be resolved in the course of further developments of this technology. Such issues are that of inventorship, i.e. who is

the inventor (can a legal entity be treated as inventor?), how to determine the person skilled in the art, decisive for assessing inventive steps, or what constitutes, for example, the relevant prior art. These issues, discussed in some detail in a conference organized by the European Patent Office in Munich in May 2018 (EPO 2018b), are the subject of many research papers (e.g. Abbott 2019; Higgins 2019).

Also in the area of copyright, if AI-controlled machines generate 'artistic' works, issues of ownership, authorship and creativity arise. Although adopted at an early stage of AI development, the UK[6] and the Irish[7] copyright laws assign the copyright in 'works generated by a computer in circumstances such that there is no human author', to 'the person by whom the arrangements necessary for the creation of the work are undertaken.' So far, no other legislator has adopted any explicit rules addressing these AI-related copyright issues. According to Gerald Spindler, at present the copyright does not protect AI as a concept or as an algorithm. In Spindler's understanding, the AI as a code is protected based on the EU Directive 2009/24/EC of 23 April 2009 on the Legal Protection of Computer Programs.[8] Moreover, if AI is adapted from a database, the structure of that database can enjoy protection under the EU Directive 96/9/EC of 11 March 1996 on the Legal Protection of Databases.[9] Spindler also foreshadows that future AI developments may well result in a situation in which it will no longer be possible to attribute the activities of AI to the 'author', which will require new legal solutions (Spindler 2019, 1050). In line with Spindler, at present the view seems to prevail that results which AI-controlled machines 'produce' are not works protected by copyright (Gervais 2019; Ginsburg and Budiardjo 2019; Ambrosio 2019).

Conclusions

Since AI has become operational in practice, the world has changed and the Earth is facing technological, ethical, socio-economic, and even political challenges. Countries and regions show a diverse level of preparedness to meet these challenges adequately. Whereas the solid figures on investment in AI developments and patenting of AI-related inventions identify China, US, Japan and Korea as the main players in the field, Europe excels particularly with plans for future investments in AI-related technologies and the existing regulations and ethical commitments. Interestingly, China, a pioneer in AI education, has plans to lead the world in AI governance through development of standards, including standards on ethics and social issues related to AI (Belton *et al.* 2019, 72). Authors who criticize the 'hands off approach' of the US, resulting in ceding the leadership to other countries, see the EU with its mandatory General Data Protection Regulation as having the first-mover advantage, and China 'aggressively writing standards for emerging technologies to benefit its own firms' (Belton *et al.* 2019, 71, 74). The problem for Europe may be that

6. Copyright, Design and Patents Act 1988 c 48, § 9 (3) in connection with § 178.
7. Copyright and Related Right Act 2000, Part I S. 2 (1) in connection with Part II S. 21 (f).
8. OJ EU No. L111/16 of 5.5.2009.
9. OJ EU No. L77/20 of 27.3.1996.

it will have difficulties in exploiting its 'first mover advantage' to have an impact on developments of international standards, without the Europe-developed AI technologies having a decisive share in the global market of AI-related technologies.

The EU Commission's optimistic claim that 'The main ingredients are there for the EU to become a leader in the AI revolution, in its own way and based on its own values' (EU Commission 2018a, 19) for the time being lacks a solid factual basis. Europe will only realize the chances that AI technology offers if it firmly stands by its commitment to invest billions of public and private funds in development, research and take-up of AI technology. This means investing in and nurturing all three essential AI elements, i.e. data, algorithms and computing power, including a competitive digital infrastructure. In all these areas Europe is in urgent need of catching up with its global competitors and stopping the increasingly looming data divide. To incentivize private investment and establish a serious European global digital player, the EU should follow the Airbus model, the only real global player originating from a European initiative. The EU should take care and actively support the still existing real 'main ingredients', such as Ericsson, Nokia, Philips, Siemens, SAP, and the many innovative start-ups, before it is too late. It should handle antitrust law with care, i.e. supporting genuine European interests. The Corona crisis has brutally demonstrated what it means to be dependent.

'Trustworthiness' is an important, essential ethical, legal and technical ramification for the beneficial practical application of AI technology and its public acceptance. It requires not only ethical and legal rules but also investment in developing and implementing the necessary algorithms. However, no matter how important trustworthiness is, it is not an end in itself, it does not solve substantive problems, but makes their solutions by AI acceptable to society.

References

Abbott R (2019) How inventive machines will change the ultimate test of patentability. *Mitteilungen der Deutschen Patentanwälte*, 377–385, 488–498.

Ambrosio Ch (2019) Unsettling robots and the future of art. *Science* **365**, 38–39.

Belton KB, Audretsch DB, Graham JD and Rupp JA (2019) Who will set the rules for smart factories? *Issues in Science and Technology* (Spring), 70–76.

Benkler Y (2019) Don't let industry write the rules for AI. *Nature* **569**, 161.

Bostrom N (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bratko I (1994) Machine learning and qualitative reasoning. *Machine Learning* **14**, 305–312.

Butler KT, Davies DW, Cartwright H, Isayev O and Walsh A (2018) Machine learning for molecular and material science. *Nature* **559**, 547–555.

Cassell J (2019) Artificial intelligence for a social world. *Issues in Science and Technology* (Summer), 29–36 (at 30).

Cyranoski D (2018) Chinese firms enter the battle for AI talent. *Nature* **563**, 260–261.

Cyranoski D (2019) AI is selecting reviewers in China. *Nature* **569**, 316–317.

DeVries JV, Dance GJX and Krolick A (2020) The secretive company that might end privacy as we know it. *The New York Times*, 18 January 2020, updated 10

February 2020, www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

Eliot T (1944) *What is a Classic? – an Address Delivered Before the Virgil Society*. London: Faber & Faber.

European Commission (2018a) *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe*. COM(2018) 237 final.

European Commission (2018b) *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan On Artificial Intelligence*. Doc COM (2018) 795 final, under F. 'Ethics by design and regulatory framework', 17.

European Commission (2020) *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. Brussels, COM(2020) 65 final, 2.

European Patent Office (2017) *Patents and the Fourth Industrial Revolution*. Munich: EPO.

European Patent Office (2018a) Guidelines for examination, G-II, 3.3.1 artificial intelligence and machine learning. Available at www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_3_1.htm (accessed 14 September 2019).

European Patent Office (2018b) *Patenting Artificial Intelligence*. Conference, May 2018, Munich: EPO.

Fleming L *et al.* (2019) Government-funded research increasingly fuels innovation. *Science* **364**, 1139–1141.

Fleming N (2018) Computer-calculated compounds: researchers are developing artificial intelligence to discover drugs. *Nature* **557**, S55–S57.

Gervais D (2019) La machine en tant qu' auteur. *Propriété Intellectuélé* **72**, 7–12.

Ginsburg JC and Budiardjo LA (2019) Authors and machines. *Berkeley Technology Law Journal* **34**, 343–448.

Gurkaynak G, Yilmaz I and Haksever G (2016) Stifling artificial intelligence: human perils. *Computer Law & Security Review* **32**, 749–758.

Higgins B (2019) The role of explainable artificial intelligence in patent law. *Intellectual Property & Technology Law Journal* **31**(3), 3–8.

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019) *The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*. Available at https://ec.europa.eu/futurium/en/ai-alliance-consultation (accessed 28 April 2020)

Ho D (2020) Artificial intelligence in cancer therapy. *Science* **367**, 982–983.

Kelly É (2020) EU makes move to ban use of facial recognition systems. *Science Business* 21 January 2020, https://sciencebusiness.net/news/eu-makes-move-ban-use-facial-recognition-systems.

Kim KG (2016) Deep learning. *HIR Health Informatics Research* **22**(4), 351–354.

Lipton B (2020) *Records on Clearview AI Reveal New Info on Police Use*. Available at www.muckrock.com/news/archives/2020/jan/18/clearview-ai-facial-recogniton-records/ (accessed 20 January 2020).

Lowe D (2018) AI designs organic syntheses. *Nature* **555**, 592–593.

Markey HT (1989) Science & law: the friendly enemies. *IDEA Journal on Law and Technology* **30**, 13–19.

Merwis J (2019) Top democrat wants big AI push. *Science* **366**, 787.

O'Meara S (2019) China's ambitious quest to lead the world in AI by 2030. *Nature* **572**, 427–428.

Organisation for Economic Co-operation and Development (OECD) (2019) *Recommendation of the Council on Artificial Intelligence*. Available at www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm (accessed 3 September 2019).

Pisano ED (2020) AI shows promise for breast cancer screening. *Nature* **577**, 35–36.

Ratner M (2018) FDA backs clinician-free AI imaging diagnostic tools. *Nature Biotechnology* **36**, 673–674.

Report (2018) FDA approves stroke-detecting AI software. *Nature Biotechnology* **36**, 290.

Rümelin JN (2019) Digital humanism. *Max Planck Research* **2**(19), 10–15.

Savelli F and Knierim JJ (2018) AI mimics brain codes for navigation. *Nature* **557**, 313–314.

Servick K (2019) Computers turn neural signals into speech. *Science* **363**, 14.

Silver D, Hubert Th, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot S, Sifre L, Kumaran Dh, Graepel Th, Lillicrap T, Simonyan K and Hassabis D (2019) A general reinforcement learning algorithm masters chess, shogi and go through self-play. *Science* **362**, 1140–1144.

Spindler G (2019) Copyright and artificial intelligence. *International Review of Intellectual Property and Competition Law* **50**, 1049–1051.

Stergiou C and Siganos D (2014) *Neural Networks*. Available at http://srii.sou.edu.ge/neural-networks.pdf (accessed 4 July 2018).

Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, MacNair CR, French S, Carfrae LA, Bloom-Ackermann Z, Tran VM, Chiappino-Pepe A, Badran AH, Andrews IW, Chory EJ, Church GM, Brown ED, Jaakkola TS, Barzilay R and Collins JJ (2020) A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.

Straus J (2017) Intellectual property rights and bioeconomy. *Journal of Intellectual Property Law & Practice* **12**, 576–590.

The Economist (2016) *The Rise of the Superstars*. Special report 'Companies'. *The Economist* 17 September.

The Economist (2018) GrAIt Expectation. *The Economist* 31 March, 3–12.

The Economist (2020) *Technology Quarterly*, Report 'Technology in China' under 'A New Trinity', *The Economist* 4 January, 9.

Thomas PhS, Castro da Silva B, Barto AG, Giguere St, Brun Y and Brunskill E (2019) Preventing undesirable behavior of intelligent machines. *Science* **366**, 999–1004.

Ullman Sh (2019) Using neuroscience to develop artificial intelligence. *Science* **363**, 692–693.

Waters R (2020) Microsoft throws weight behind data movement. *Financial Times* 21 April, 5.

Wong D and Yip S (2018) Machine learning classifies cancer. *Nature* **555**, 446–447.

World Intellectual Property Organization (WIPO) (2019) *Technology Trends 2019 Artificial Intelligence*. Geneva: WIPO.

About the Author

Joseph Straus, Dr. jur., Dres. jur. h.c., is Professor of Law at the Universities of Munich and Ljubljana and Emeritus Director of the Max Planck Institute for Innovation and Competition, Munich. He is a Foreign member of Accademia Nazionale dei Lincei, Corresponding Member of the Slovenian Academy of Sciences

and Arts, Foreign Member of the Academy of Sciences and Arts of Bosnia and Herzegovina, Member of Academia Europaea (Chair of the Law Section 2009–2015). He was the Chair of the Intellectual Property Rights Committee of the Human Genome Organization (HUGO), 2009–2015 and Chair of the Permanent Working Group on Intellectual Property Rights of The Federation of European Academies of Sciences and Humanities (ALLEA), 2009–2019.