Annual Reviews in Control 53 (2022) 391–404

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Review Article

From ethics to standards – A path via responsible AI to cyber-physical production systems

István Mezgár [a], József Váncza [a,b,*]

[a] Institute for Computer Science and Control, Eötvös Loránd Research Network, Budapest, Hungary
[b] Department of Manufacturing Science and Technology, Budapest University of Technology and Economics, Hungary

ARTICLE INFO

Keywords:
Artificial intelligence
Cyber-physical production system
Agents
Ethics
Control
Trust

ABSTRACT

The central claim of the paper is that the development and control of Cyber-Physical Production Systems (CPPS) requires a systematic approach to handle and include explicit ethical considerations. Since the contribution of artificial intelligence (AI) technologies, and of agent-based models in particular, was instrumental in the evolution of CPPSs, approaches of ethical AI should be endorsed in CPPS development by design. The paper discusses recent advances for ethical AI and suggests a pathway from ethical norms towards standards. As it is argued, taking the responsible AI approach is promising when tackling the main ethic-related challenges of Cyber-Physical Production Systems. We expose a number of dilemmas to be resolved so that AI systems incorporated in CPPS cause no damages either in humans, equipment or in the environment and increase the trust in the users of current and future AI technologies.

1. Introduction

The fast evolution of economies demands new information technologies to facilitate increased efficiency, responsiveness, and more recently, also sustainability. Digitalisation can offer solutions with its various new technologies. Artificial Intelligence (AI) is identified as the basic technology that when combined with other emerging technologies can effectively multiply their performance in a great extent (like that of the Internet of Things, Big Data, or more recently, of 5/6 G). Additionally, AI can substitute or provide intensive help for human decision-makers and workers alike, so AI can be applied in all sectors of the economy providing strategic advantages. Indeed, Gartner states that AI will become the biggest megatrend of the next decade (Panetta, 2018). The introduction and proliferation of AI technologies was instrumental also in the development of Cyber-Physical Production Systems (CPPS) (Monostori et al., 2016) which is in the focus of this paper.

In the mainstream AI research, and in applications particularly, the concept of agents started to make a break-through three decades ago or so when the emphasis shifted from goal-orientated, logic-centred to utility driven, rational behaviour. By definition, agents act in an environment, where after making observations they change the environment with their actions in a way which serves best their own interest. What really matters is that an agent does the "right thing", even with bounded computational resources (Russell & Wefald, 1991). This concept of AI intensified further research in machine learning (when the performance of an agent can be improved on the basis of its accumulated experience), in multi-agent systems (when the environment is populated also by other agents), and, more recently, also in robotics (when human and machine agents can be united in a collaborative system) (Kemény, Váncza, Wang, & Wang, 2021).

The agent-based concept of AI was early welcome also in production, at all its levels, from the control of production equipment up to the management of supply chains and global production networks. No doubt, agents held the promise of some much-sought properties like autonomy, responsiveness, and cooperation (Monostori, Váncza & Kumara, 2006). In our view, the agent concept was one of the key enabling technologies of CPPS. In particular, agents contributed to the development of CPPSs whose main traits are (1) intelligence or smartness of elements, (2) connectedness that enables harnessing the data/-knowledge and services available in their environment, as well as (3) responsiveness, an ongoing interplay between the physical system entities and their representations (Monostori et al., 2016).

However, really autonomous agents still have a long and difficult path for general industrial acceptance. Any real-life application must comply with a number of variegated but strict requirements in terms of correctness, reliability, robustness, and most importantly, safety. But

* Corresponding author.
E-mail address: vancza@sztaki.hu (J. Váncza).

https://doi.org/10.1016/j.arcontrol.2022.04.002
Received 10 January 2022; Received in revised form 21 March 2022; Accepted 5 April 2022
Available online 16 April 2022

how can a correct and safe operation come along with autonomy? Can we give any warranties that an autonomous system complies all the time during its operation with the purpose(s) which motivated its creation? When interacting with humans, can autonomous systems honour and give priority to human values, in any case and under any circumstances? Can they resolve conflicts between different value systems, as it happens routinely in human decision-making? Has CPPS any advantage when trying to find some resolution to the dilemma of competitive, and, at the same time, sustainable manufacturing? All these questions stretch far beyond the realms of information, communication and production technologies and raise moral and ethical issues as well.

Furthermore, in production there is also a paradigm with direct relations and impact to economy and society, the recently emerging biological transformation of manufacturing (BTM) (Byrne et al., 2021; Byrne, Dimitrov, Monostori, Teti & van Houten, 2018). This is about the "use and integration of biological and bio-inspired principles, materials, functions, structures and resources for intelligent and sustainable manufacturing technologies and systems with the aim of achieving their full potential." One can take BTM as a systematic application of our knowledge of biological processes to the optimisation of production, from product design up to elaborating the framework of sustainable manufacturing (Neugebauer, Ihlenfeldt, Schliesman, Hellmich & Noack, 2019). In the field of product development, the virus-built batteries, protein-based water filters, cancer-detecting nanoparticles can be listed as examples of future products (Hockfield, 2019). These developments coined as living manufacturing systems borrowed many concepts from AI research, as an analogy drawn between systems of life and of manufacturing shows (Monostori & Váncza, 2020). Clearly, BTM raises also new ethical issues for production engineering.

AI theories, technologies and applications evolved so fast in the last two decades that leading AI researchers realized that rather sooner than later we have to face and resolve not only technical, but deep social, ethical and legal problems connected to AI. The intensive research on AI and other ethics-sensitive technologies resulted in the formation of a number of strategies, guidelines, proposals for ethical and trustworthy AI applications. Some of these efforts have generated already the first wave of laws, regulations and technical standards (Schmelzer, 2020). No doubt, AI technologies dominantly increase the efficiency, agility, and adaptability of production too. The significance of the already existing and future AI-related regulations and standards cannot be underestimated, but it is also clear that further research is needed, even in the narrowed scope of CPPS, to facilitate the development of trustworthy and ethical systems in manufacturing (Mezgár, 2021).

The focus of the paper is set on how ethical aspects can be involved into the design and operation of a CPPS through regulations and standards under development. Hence, it does not deal in depth with the philosophical relations of ethics, neither with the ethics connected to autonomous vehicles, health and military applications. The structure of the paper is as follows: Section 2 discusses AI, autonomous agents and their relation to ethics and trust. Section 3 presents the relations of ethics, law and standards, and surveys existing standards in connection with ethics and trustworthy AI applications. Section 4 introduces shortly the components and principles of ethical, trustworthy AI software development. Section 5 describes how ethical aspects can be taken into consideration during the design of AI applications in any CPPS environment, giving also some application examples. Section 6 discusses several dilemmas and summarizes recommendations on the ethically conscious design and development of cyber-physical production systems. Finally, Section 7concludes the paper.

2. Artificial intelligence and its relations to ethics and trust

2.1. Types of artificial intelligence

Artificial intelligence has many different definitions according to the time, focus, and goal of the actual accent. A general, broad definition of AI has been proposed by European Commission (EC, 2019): "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 behaviour by analysing how the environment is affected by their previous actions." The study takes AI as a scientific disciple with its main methods like machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization) and learning (of which deep learning and reinforcement learning are specific examples), robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems) and language processing. Note that there is a trend to narrow the scope of AI to data-driven approaches. As a typical example, (Theodorou & Dignum, 2020) says that "Artificial Intelligence technologies are able to process vast amounts of data and to infer patterns or even to draw conclusions from that data. The algorithms enabling these human-like cognitive processes of making predictions or decisions can improve by themselves through experience and use of data and often rely on machine learning and neural networks". In what follows the more inclusive – and more traditional – approach to AI is taken.

In general, there can be distinguished three basic categories of AI with regard to its relation to human intelligence and faculties.

- The "weak" or artificial "narrow" intelligence (ANI) focuses on some specific task to do well, but it doesn't have the ambition to substitute human intelligence. It is rather aimed at reinforcing human cognitive abilities, like a hydraulic press increases the impact of muscular force. Hence, it re-constructs, simulates and supports human decision-making with respect to some specific task.
- The artificial general (or "strong") intelligence (AGI) has the ambition to match human level intelligence, principally in any fields and types of human activity. Today such AGI systems do not exist yet.
- The third category of AI is the artificial super-intelligence (ASI) that would exceed human intelligence and faculties, so that even the best human minds cannot compete it. The point where the trajectory of artificial intelligence would reach human-level intelligence is called the point of singularity beyond of which ASI systems could self-improve and build even more intelligent systems with hardly predictably behaviour (Müller, 2020). According to another approach (Kurzweil, 2005) humans in the future will combine biological and non-biological intelligence, with a more and more dominating non-biological component. Developments of neuromorphic computing (Monroe, 2014) and brain robotics (Zhang et al., 2020) point in this direction.

In any case, in order to cooperate with different types of AI systems and to gain the confidence and trust of their users, ethical measures have to be defined that make possible the symbiosis of human and AI systems.

2.2. AI in the view of the public

As the application of AI proliferated literally in every domain of human activities, it became more and more important to assess how the society reacts to the above developments, to the introduction and day-to-day use of AI applications, to the risks one can anticipate or only feel. The Oxford Commission on AI analysed the global opinions on using AI in decision-making based on a broad sample of 154,195 respondents in 142 countries (Neudert, Knuutila & Howard, 2020). According to this study, automated decision-making AI is considered potentially harmful with highest rate in North America (47%) and definitely smaller in East Asia (11%) see Table 1. The data for the study

Table 1
Global risk perception of AI decision making in %, by region (Neudert et al., 2020).

Region	Mostly harm	Mostly help	Neither	Do not know
Latin America & Caribbean	49	26	19	6
North America	47	41	12	0
Europe	43	38	15	5
Central Asia	34	36	17	13
Middle East	33	38	19	10
South Asia	33	31	17	19
Africa	31	41	16	12
Southeast Asia	25	37	21	17
East Asia	11	59	12	18

was collected from the 2019 World Risk Poll (WRP), and at least 1000 respondents were surveyed in each country. The probability-based samplings represented the entire population aged over fifteen. The sentiments on AI were compared across gender, education, individual-level income, employment hours, attitude risk and other variables, all detailed in the study.

Another recent study (Gillespie, Lockey & Curtis, 2021) surveying opinions in the USA, Canada, Germany, UK and Australia states that 28% of citizens are willing to trust AI systems in general. Two out of five citizens are unwilling to share their information or data with an AI system and a third are unwilling to trust the output of AI systems. For methodology, statistical procedures, and demographic details, see Gillespie et al. (2021). More recently, Ipsos published a study on trusting companies that use artificial intelligence (Ipsos, 2022). The results which show a clear difference between high-income and emerging countries in attitudes towards AI are based on a 28-country survey using Ipsos's Global Advisor online platform. Citizens from emerging countries are significantly more likely than those from more economically developed countries to report being knowledgeable about AI, to trust companies that use AI, and to have a positive opinion on the impact of AI-based products and services. The likelihood to trust companies that use AI is highest amongst business decision-makers (62%), business owners (61%), with a higher-education degree (56%), and lowest amongst those who are 50 or older (44%), with no higher education (45%). The global country average is 50%. Ipsos interviewed a total of 19,504 adults aged 18–74 in the United States, Canada, Malaysia, South Africa, and Turkey, and 16–74 in 23 other markets between November 19 and December 3, 2021. The sample consists of approximately 1000 individuals in each country. An interesting news on the expected effect of the wide-spread and important influence of AI on all segments of the society is that the Pope himself asked for prayers for AI and robotics in 2020: "We pray that the progress of robotics and artificial intelligence may always serve humankind" (Pope Francis, 2020).

2.3. The relation of ethics and AI

Ethics has been studied by many researchers from different disciplines, so ethics and the connected moral have a number of different approaches, definitions and taxonomies. A short definition of ethics and moral are the following "Ethics is a set of beliefs that a society conveys to its individual members, to encourage them to engage in positive-sum interactions and to avoid negative-sum interactions" (Kuipers, 2020), while "moral is relating to the accepted good or bad behaviour, fairness, honesty, etc. that each person believes in, rather than to laws" (Cambridge Dictionary, 2021).

In this section a technical approach will be applied focusing on possible relations to CPPS, so extended definitions are introduced of the two terms. A more descriptive definition of ethics can be given as ethics is a set of moral principles, standards (or form of conduct) provided by external sources (certain community or social setting) to which an individual belongs, distinguishing the difference between "good and bad "or "right and wrong". Moral refers to an individual's own internal principles regarding "right and wrong", so there are personal principles created and upheld by the individuals themselves. Ethics and moral both are influenced e.g., by culture, society, geographical position, profession or field of application.

Ethics can be described from different aspects; in the followings a short overview is given on applied ethics that defines what a person can do in a specific situation on a particular field of action in real-life situations. Applied ethics has many specific areas, such as machine ethics, ethics of technology (techno-ethics), cyber-, digital ethics. In case of artificial intelligence, the ethics of AI defines the ethical and moral obligations and duties both of an AI system and its developers (Keng and Wang, 2020).

New technologies offer new possibilities, more physical and mental power that can provide new chances that were not present before. This new behaviour can exceed the existing ethical boundaries, hence time and again, the ethical (legal system) concepts have to be renewed (EC, 2018a), (EC, 2018b).

2.4. Types of ethics

While analysing the connection of ethics of humans and the quickly developing technologies of AI and autonomous systems, three main fields can be distinguished that are partially overlapping each other; the human ethics, the human-machine ethics, and the machine/autonomous system ethics.

Ethics basically studies the human relations and issues emerging in a society. It helps defining some protective boundaries around people who decide and act autonomously as members of a community. There is an essential need to delineate right and wrong decisions, to distinguish good and bad outcomes. The crux of the problem is that rules and constraints should precede cost-benefit calculations or just the other way around, decisions and actions should be judged primarily on the basis of their perceived impact. As an unresolved problem, it has in the recent philosophical discourse two basic approaches (Nagel, 2021).

The so-called consequentialist justification evaluates and weights actions in terms of their long-term impact, their implied costs and benefits (Anscombe, 1969). The outcome matters only, whether it is good or bad, better or worse. The application of this principle requires advanced perceptual observation, a deep knowledge of the causal relations of the world, as well as the employment of an extremely efficient logical reasoning mechanism. It is rightly debated that this basically utilitarian approach can be untenable whenever our situation assessment or knowledge is burdened by uncertainties, or time-critical decisions are to be made. Note that the original AI concept of rational agents fits this stance to ethics.

The alternative way suggests sticking to some basic principles that embody the notion of values in a society. The so-called deontological approach directly evaluates the rightness or wrongness of actions and policies, without considering explicitly their direct impact. These rules safeguard the integrity of a society by blocking actions that would lead to greater evil and promote actions that would produce greater good. Whether acts embody intrinsic values, or rules related to their employment are merely shortcuts (Kahneman, 2011) that make social interaction efficient (or possible at all) are broadly debated, just like the origin and evolution of deontological principles.

There is clearly a tension between the two approaches and even common sense intuition can create dilemmas when a resolution is hard to find. The so-called reflective equilibrium still tries to find a balance: departing from general principles it makes considered value judgements about specific situations and outcomes, and adjusts both iteratively till a consolidation is found (Brandt, 1990). Hence, deontological intuition which captures a kind of moral minimum can be overridden at times, if anticipated consequences of actions provide strong enough arguments to do so. As it seems from the contemporary discussions about moral

intuitions, this pragmatic approach is getting more and more momentum (Nagel, 2021).

2.5. Ethics and trust

Ethics and trust are in close reciprocal connection. High trust environment encourages to build better ethics, while ethical behaviour support trust building. Trust can be defined as a psychological condition comprising the trustor's intention to accept vulnerability based upon positive expectations of the trustee's intentions or behaviour (Rousseau, Sitkin, Burt & Camerer, 1998). Trust can cause or result from trusting behaviour (e.g., cooperation, taking a risk), it can motivate an action, but is not behaviour itself.

Specifically, trust in the technology is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004). In case of socio-technical systems trust appears in different forms. According to Luhmann, 1979 four types of trust can be differentiated: (1) intrapersonal trust, a kind of self-confidence basic trust, (2) interpersonal trust, an expectation based on cognitive and affective evaluation of the partners, (3) system trust in depersonalised systems, let it be legal, or technical, and (4) object trust in non-social objects.

In case of humans the feeling of being safe and secure is determined by the user's sense of control in an interactive system. The more a user feels in control of an interactive program, the more the user will trust the site, the program, the application. An interactive system that allows the user to feel in control should match four demands: the system is comprehensible, predictable, flexible and adaptable. This control has technical and emotional parameters, as Xu, Le, Deitermann & Montague (2014) gave a detailed analysis of user's trust building in technology.

Trust building and the ways to maintain the needed characteristics of trust in the case of autonomous systems is still in research phase. There are studies to define their components, main behaviour characteristics but the problem is very complex (Sifakis, 2019). Trust is an important component of the AI ethical frameworks (see Section 4.) as only following some ethical principles can one develop trust in AI-based autonomous systems.

The European Consumer Organisation (BEUC) canvassed more than 10.000 consumers in 9 European countries about their fears over transparency, accountability, equity in decision-making, and the management of personal data in AI-based applications (BEUC, 2020). The findings were that a large majority of responders feel that AI can be useful but most of them do not trust the effective control of the technology (an average 56%) and feel that current regulations would not protect them (about 80%) from the negative results it can cause.

2.6. Autonomous agents, responsibility and ethics

AI developments focusing on autonomous agents have driven research towards so-called ethical (or moral) agents. Svegliato, Nashed and Zilberstein (2020) proposes to make a distinction between an agent's decision-making (basically, amoral) mechanism and its moral, prescriptive model which defines its actual ethical context. Any policies of the decision-making model should be evaluated and ranked in the ethical context, and in any case the best-ranked option should be selected for action. Autonomous systems not capable of generating such decisions in the problem domain which do not pass the ethical test are considered unrealizable. Incorporating of such ethical provisions into the agent architectures is also suggested by Scheutz (2017), and, in a similar vein, the concept of ethical controller is suggested by (Trentesaux & Karnouskos, 2021).

Hooker and Kim (2019) goes further by suggesting that truly autonomous agents can be but ethical (or cannot be unethical) by design. Rooting the concept of autonomy in the deontological tradition of ethics, it is argued that ethics provides internal motivations and constraints to autonomy. An agent equipped with sufficient observation and logical reasoning faculties should be able to deduce if its actions would cross other agents' intentions and plans and violate this way their autonomy. Avoiding anything it would not like to face itself, an autonomous agent would refrain from breaching promises, breaking contracts, or risking safety of others (including humans). An important point of this model is that it distinguishes simple behaviour which could be the results of a control decision prompted by a cause from autonomous action which should have reason(s) and explicit explanation.

The above, so-called top-down approaches to ethical agents attempt to set up a clear-cut, principle-based computational model without considering practical details. Most importantly, reasoning in such a refined way puts extreme burden on computational resources, when one can hardly give any warranties on reasonable response time (let alone prompt actions). The alternative way of engineering moral agents leads via experience and learning. However, this bottom-up approach (Fisher, List, Slavkovik & Winfield, 2016; (Wallach and Vallor, 2020)) has also its risks, because any mechanism of learning implies (many) cycles of failures.

Recently, Russel (2016) proposed to combine the principle-(constraint) and training-based approaches by suggesting three core principles for designing artificially intelligent autonomous systems. (1) The machine agent's purpose must be to maximize the realization of human values, (2) even if initially it is uncertain about those values. However, (3) the machine must be able to learn about human values by observing the choices of humans in its environment. Hence, this proposition "rationalizes" agent behaviour by extending it to a faculty which learns and adheres to human values, among others those incorporated by ethics. Note that this approach has essential features in common with the reflective equilibrium proposal discussed in Section 2.4 above.

3. Relations of ethics, law and standardization

3.1. Need for legal system for AI technologies

The AI scientific community called the attentions of governments to the risks of AI systems and the approaching singularity point, so states and governments realized that new AI applications can cause problems for the whole economy and society as well. Based on the dialogue of government officials and research community AI strategies have been developed by most of the countries (OECD, 2019). These contain proposals for new AI related laws and development strategies, and frameworks for AI system developments as well (U.S. National Science, 2016; China – AI Standardization, 2020; Cihon, 2019; OECD, 2019).

Artificial intelligence ethics elevated to the top of policy agendas for governments and other stakeholder groups at both national and international levels. It is important to react in time to the risks generated by AI developments and applications, many of which raise also public concern (see Section 2.2). Legal systems and standards play vital role in creating a defensive framework for the society from the perceived, anticipated and real threats of AI, help develop and maintain trust in AI systems. Standards can add direct and detailed instructions for developers how to create ethical AI applications warranting interoperability, safety, transparency and security. Standards developed by international bodies such as ISO and IEEE can support the global control of AI development (Cihon, 2019). A properly configured legal system can raise user trust in the technology and support the development of ethical AI applications as well. There is a common understanding that their formation should go hand-in-hand with research and development.

Fig. 1. below gives general overview of the process from AI problem recognition to AI standards development. The figure shows the main steps of and the links between the legislative and standardisation processes. The stakeholders in general belong to academia, technical communities, business participants, civil society, intergovernmental organisations, and trade unions. Here we focus on the role of the three most important stakeholders:

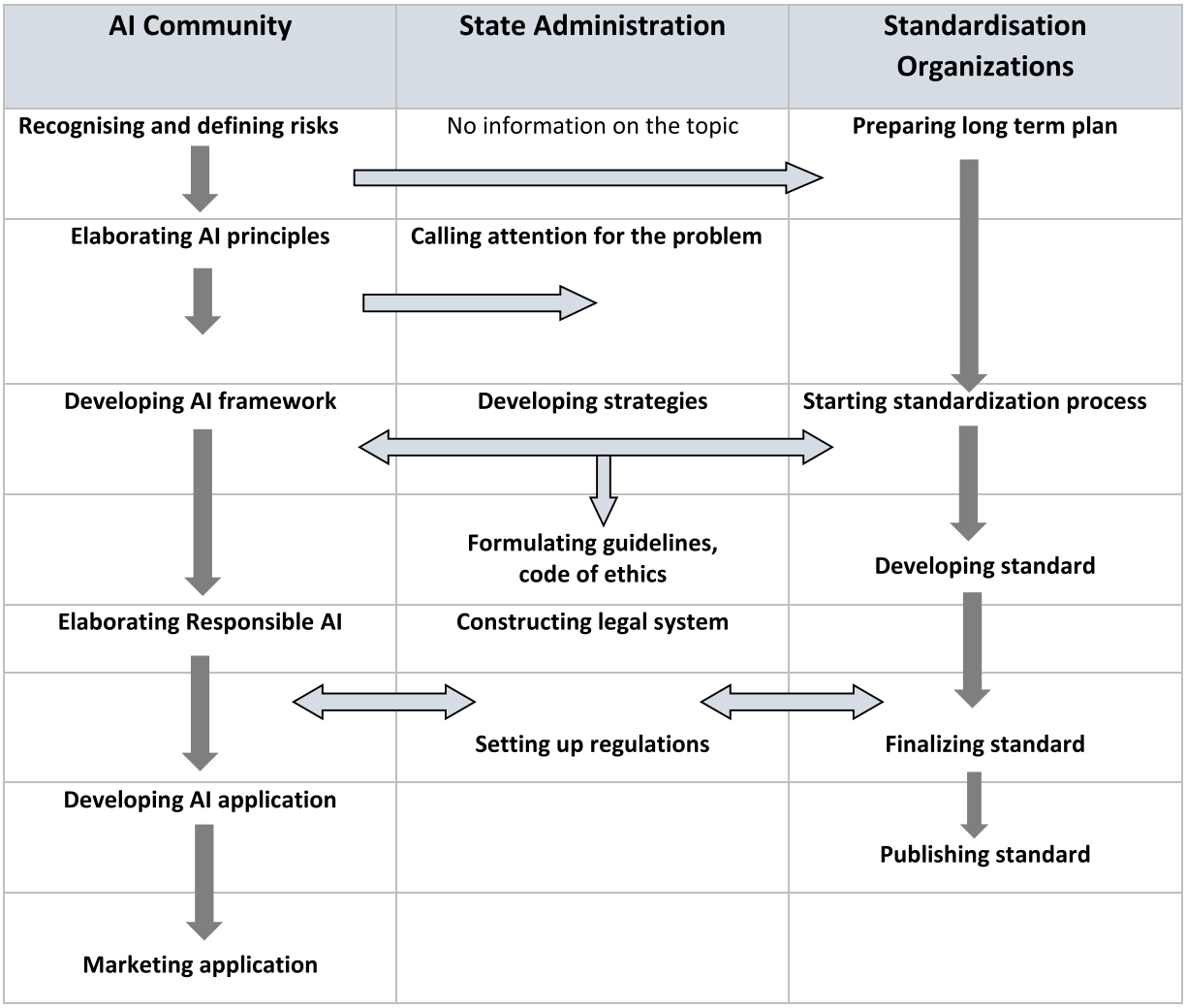

Fig. 1. Draft process from AI problem recognition to AI standards development with stakeholders' involvement.

- *AI community* – leads the innovative research and development process, always well before the other two processes.
- *State administration* – a multilevel organisation, with complex decision-making mechanisms on a number of different levels, operates relatively slowly, always relying on external information and advice.
- *Standardization bodies* – both international, national – working in long processes from the first ideas to elaborating and submitting standards, while getting through many compromises.

The selected stakeholders use different concepts, terms, expressions and wording – indeed, different languages – to elaborate their ideas. Their parallel processes have also distinct characteristics in terms of time scale, participants, negotiation methods, cultural background and the existing legal and administrative systems. As for the legislative systems, consider only two examples: in the USA 15 steps in two parallel linear processes including multiple balloting are needed while a bill (idea in a legal form) becomes a law (US Government, 2022). On the other hand, the Swedish legislation process is typically circular: starting in committees, then preparation of government draft proposal, finally parliament voting (Ministry of Justice, 2016).

The standardisation process depends on the technical environment, the initiative organization and the background of the actual standardisation organization. It is a repeatable, documented, proven process. As an example, the IEEE Standardisation Association's process has seven steps. An idea for a standard emerges; the proposed standard is developed; an IEEE sponsor conducts a vote; governing committees approve the standard; the standard is published; and it goes into a maintenance phase. Ideas for standards can arise in a variety of ways, e.g., suppliers of peripherals may realize they can save costs if an interoperability standard will be used by all, or a new and promising technology could become widely adopted/accepted if it is standardized (EDN, 2014).

As shown in Fig. 1., risk detection and evaluation in AI research and implementation is the most important first step. After the initial findings by the researchers this information is forwarded to the standardization bodies and the attention of state administration is also raised to start the appropriate standardization and law-making process. In the next phases the research community elaborates proper frameworks to structure risks and countermeasures to eliminate or mitigate the threats. Parallel with these activities continuous, step by step dialogues are going on with the state administration and the standardisation bodies. The outputs of this collaborative efforts are the proper laws and standards that match the (nearly) latest result of the AI research.

Recently, the Organisation for Economic Co-operation and Development (OECD) has published AI related guidelines called "Recommendations" which focus on how governments and other actors can shape a human-centric approach to trustworthy AI. This study identifies five complementary value-based principles for the responsible control of trustworthy AI and calls on AI developers and users to promote and implement them. In addition to and consistent with these value-based principles, the Recommendations also provide five suggestions to

policy-makers pertaining to national policies and international co-operation for trustworthy AI (Yeung, 2020). The OECD also provides a live repository that contains more than 700 AI policy initiatives from 60 countries, territories and the EU. From this site the strategies, laws and standards can be downloaded (OECD, 2021). Furthermore, the status of AI related laws, regulations and standards in the USA, UK and the EU are summarised and analysed from a lawyer's perspective in Gibson, (2022). This study covers issues of ethics too. A research firm published a report on "Worldwide AI Laws and Regulations 2020″ (Schmelzer, 2020) in which the latest legal and regulatory measures of countries worldwide are summarized in seven different areas, such as autonomous vehicles, facial recognition and computer vision, AI supported decision making, AI related data privacy, AI ethics and bias.

3.2. Guidelines and code of ethics

These guidelines are important as they define the directions of creating laws, regulations and standards in the future. Based on these expectations the tools for ethical AI applications can be defined by bringing together the technical possibilities, the directions of legal systems and the expectations/requirements of the society. The motivation for composition of these guidelines is manifold; declaration of the positive approach to AI for the society, providing guidelines for developers, and demonstration to have a good position in the competitions of AI technology for future applications.

To bridge the gap between ethics and standards new approaches (such as trustworthy, explainable, or responsible AI) have been developed that support the research and development works in this direction. Numerous ethical guidelines have been published recently; in the followings a few of them will be summarized shortly.

3.2.1. EU regulation on artificial intelligence

The EC "Proposal for a Regulation laying down harmonised rules on artificial intelligence" (EC, 2021a) presents the first ever legal framework on AI, which addresses the risks of AI and supports the development of secure, trustworthy and ethical artificial intelligence. The legal framework has a life-cycle approach, and classifies AI applications according to their risk-level representing functions as follows:

- *Unacceptable risk* is implied by AI systems or applications which are considered a clear threat to the safety, or manipulate human behaviour to circumvent the users' free will.
- *High risk* AI systems are of the following subtypes:
  - Critical infrastructures can put the life and health of citizens at risk.
  - They can be safety components of mission-critical products, such as AI application in robot-assisted surgery.
  - All remote biometric identification and classification systems are considered high-risk and subject to strict requirements.
- *Limited risk* AI systems have specific transparency obligations.
- *Minimal risk* AI systems – the majority – represent only minimal or no risk for citizens' rights or safety.

A detailed analysis on the role of standards in the EC AI regulation is given by McFadden, Jones, Taylor and Osborn, (2021). An additional regulation has been issued parallel with (EC, 2021a), the "European approach to new machinery products" (EC, 2021b). Machinery products cover a wide range of professional and consumer products, from e.g., robots to 3D printers, industrial production lines. The new Machinery Regulation will ensure that the new generation of machinery guarantees the safety of users and consumers and encourages innovation. While the AI Regulation deals with the safety risks of AI systems, the Machinery Regulation focuses on the safe integration of the AI system into the overall machinery.

3.2.2. China - ethical norms

In September 2021, the National Governance Committee for the New Generation Artificial Intelligence published the "Ethical Norms for the New Generation Artificial Intelligence" (NGCC, 2021). It aims at integrating ethics into the entire life-cycle of AI, to provide ethical guidelines for natural persons, legal persons, and other related organizations engaged in AI-related activities. This set of norms is formulated in order to refine and implement the "Governance Principles for the New Generation Artificial Intelligence", to enhance the ethical awareness on artificial intelligence and the behavioural awareness of the entire society, to actively guide the responsible AI research, development, and application activities, and to promote healthy development of AI. According to experts many aspects of the Chinese principles seem similar to those of the EU, both promoting fairness, robustness, privacy, safety and transparency. Their prescribed methodologies however reveal clear cultural differences.

3.2.3. Russia – code of ethics

A code of ethics of artificial intelligence has been issued in October 2021 by the AI Alliance, jointly with the Analytical Centre Russia. The Code will become part of the Artificial Intelligence federal project and the "Strategy for the Development of the Information Society for 2017–2030″ (Nocetti, 2020). It establishes general ethical principles and standards of conduct to guide those involved in activities using artificial intelligence. The Code applies to relations involving ethical aspects of the creation (design, construction and piloting), implementation and use of AI technologies at all stages of the life-cycle, which are currently not regulated by Russian law or other regulatory acts. Joining the code is voluntary.

3.2.4. USA - ethics for the intelligence community artificial intelligence principles and framework

The Intelligence Community (IC) released the "Principles of Artificial Intelligence (AI) Ethics for the Intelligence Community" and the related "Artificial Intelligence Ethics Framework for the Intelligence Community". These principles and framework will guide the IC's ethical development and use of AI (IC, 2020). The "Principles of AI Ethics" demonstrate the IC's commitment to ensuring its use and implementation of AI respect the law, protect privacy and civil rights, are transparent and accountable, remain objective and equitable, appropriately incorporate human judgement, are secure and resilient by design, and incorporate the best practices of the science and technology communities. IC data scientists, privacy and civil right officers and other key stakeholders collaboratively developed the AI Ethics Framework to ensure the incorporation of the basic Principles of AI Ethics during the complete life-cycle of products and services.

The above policies are in common that the basic ethical principles and the life-cycle approach can be found in each one. At present keeping all the suggestions is voluntary. All in all, over 60 countries issued different AI strategies and ethical guidelines (OECD, 2021). Their analysis is beyond the aims of this paper. As the European proposal is the first real comprehensive material in the field, many discussions and polemics were published in connection with this act (Veale & Borgesius, 2021).

The differences between the guidelines originate from the distinct cultural, religious, social and economic background of the countries. There are proposals based on a centralized social system, others are more connected to the central state. The European proposal puts the citizens' rights in the centre (e.g., privacy). In the USA the private business sector has strong influence on the suggested principles, so it is more liberal, competitive, and allows more freedom for the AI product developing companies. The most critical point of the European proposal is the high-risk qualification of all remote biometric identification and classification systems, as they are subject to strict requirements according to the proposal.

3.3. Relations of law, regulations and standards

3.3.1. Definitions and relations

It is important to make clear the differences between law, regulation and standard. The differences lie in the origin, function and validity of these rule systems. Their extended and refined definitions are given in the followings based on (Cambridge Dictionary, 2021), their short definitions and main characteristics are introduced and compared in Table 2.

- *Law* is a rule, usually made by a government, which is used to order the way in which a society behaves. The legal system is the system of rules of a particular country, group, or area of activity that everyone in a country or society must follow to be legal.
- *Regulation* is an official rule or the act of controlling something made by a government or some other authority. Detailed instructions on how laws are to be carried out (implementation of law) and are sometimes referred to as "rules" or "administrative laws", their application is mandatory.
- The *technical standard* is an official rule, unit of measurement, or way of operating that is used in a particular area of manufacturing or services. Standards are providing specifications (guidelines or requirements) for products, services or systems. If used consistently, they ensure quality, safety and efficiency as well. Conformity with standards is voluntary.

Today, standards have strategic importance as in the digital era networked, distributed systems have to collaborate; data and information (knowledge) have to be exchange in a safe, reliable, secure and exact mode. Standards provide advantages in the research, development and operation phases and on the market as well by the easy change of parts or connecting different software modules in a reliable way by standardised protocols. Additionally, common standards and regulations will be necessary to ensure the safety of users; this would provide consumers trust in new innovative solutions and decrease potential health, safety, security and privacy risks (Zachariadis, 2019). If one can dominate the standardization process, then it can have direct business advantages as well (Ding, 2018).

3.3.2. Connection between ethics, laws and standards

Ethics and moral can be handled as a mental, philosophical category, providing freedom for people controlled "only" by internal believe or the expectations of other people. Laws, regulations and standards are generated by groups of people, they set administrative limitations, or give recommendations. The bridge that connects these two different sets of controlling barriers is trust.

Good laws, regulations generate trust, feeling of safety in the people, support to keep ethical, moral – even not written – standards and norms. High-level moral, ethical environment (such as company ethical rules and guidelines) can generate trust to other people, motivate to keep regulations, to act according to standardised processes. On the other hand, standards provide guidance, feeling to be safe as acting according to a matured process, even though users can have the impression of own control over a system. Hence, using a standard-based, well designed and tested equipment, systems – including software – will generate trust as well. Clearly, there is some kind of reciprocity, indirect or collateral relation between ethics and standards.

According to a recent international analysis (Gillespie et al., 2021), 66–79% of citizens of the involved countries believe the impact of AI on society is uncertain and unpredictable, and their overwhelming majority (96%) require that AI governance challenges be carefully managed. More than 57% would be more willing to use AI systems if assurance mechanisms were available such as AI ethics certifications, national standards for transparency, and AI codes of conduct.

Social or professional communities, closed groups of people – like also companies – typically form their own ethical norms, guidelines. Investigating the evolution of such systems can help understand how ethics, law and standards interplay and evolve together. It is a common experience that explicit ethical norms in a company generate positive feedback: employees feel better, more safe in the workplace, and have greater trust in the company as the company can trust them better as well (Rothenberger, Fabian & Arunov, 2019). Nowadays, there are already some companies that build into their guidelines long-term ethical AI considerations too, as it is the current practice of a leading CPPS technology provider (Bosch, 2020). Their engineers participate in AI courses, studies, and standard developments as well so as to be able to design products with embedded ethical AI. These are accepted and trusted better also by their users. Based on the collected experiences the ethical guidelines can be continuously evolved generating in this way higher trust both in their workers and their products' users.

3.3.3. Technical standards for AI ethics – examples

AI is already in widespread use today, but related technical standardisation is still in its infancy. However, there are numerous existing standards that relate to building ethical components – e.g., standards for security (GDPR), cobot safety (ISO/TS 15,066) – that can be applied in the development of trustworthy AI systems. A good source for AI policies, initiatives and strategies of governments, and technical organisations is the collection of OECD (OECD, 2019) which list and compares international and domestic AI standards. Below, we highlight those which focus on ethical issues and can directly be applied in CPPS system development.

- International organizations for standardisation such as ISO and IEC deal with ethical aspects of AI. For the time being they have published six standards, and 22 are under development (ISO, 2021). Of special interest is the "Overview of trustworthiness in artificial intelligence" (ISO, 2020).

Table 2
Main characteristics of regulating concepts (adapted from (Mezgár, 2021)).

Aspect, category	Ethics	Moral	Law	Regulation	Technical standard
Short definition	System of rules of conduct – "right" and "wrong" defined by the society.	Individual's personal principles on "good" and "bad".	System of rules defined by a government or community.	Implementation of law for a narrower field.	Technical specification for a product, system or service initialized by companies.
Origin	External from a social system.	Internal from the individual.	Official high-level administration of a country.	Lower-level official administration of a community.	Professional recognized body, organization.
Reason to keep it	Not to be shamed by others/ society, match the expected behaviour.	Internal believe, self-motivation to keep them.	Mandatory, not keeping it results penalty, can generate trust.	Mandatory like laws.	Conformity with standards is voluntary, supports interoperability, raise safety, efficiency and quality.
Changing in time	Slow changes as the society moves on.	In case the individual's beliefs change.	According to the needs of society.	Periodically, demands from the users, developers.	According to the technical development, needs of industry.
Influenced	Connected to a particular time and place.	Cultural, religion background.	Society, existing legal - political background.	Local, application, sector level.	Professional aspects, technical level.

- Technical communities, and eminently, IEEE put much emphasis on standardisation issues. IEEE launches the "Global Initiative on Ethics of Autonomous and Intelligent Systems" project on AI standard development (IEEE, 2021). The IEEE P70xx is a series of standards on "Ethics of Autonomous and Intelligent Systems". Furthermore, other 13 related standards were partially issued or are under development.
- The EU standards bodies CEN and CENELEC officially created a Focus Group on AI, in support of ISO/IEC SC42. Standard organisations like ETSI and CENELEC have published agendas for AI standardisation, partially motivated by the proposed EU AI regulation's framework for standards. ETSI has focused on security issues surrounding AI and machine learning, while CENELEC has a strong focus on trustworthiness and ethics.
- In the US, National Institute of Standards and Technology (NIST) has defined that trustworthiness standards should include guidance and requirements for accuracy, explainability, resiliency, safety, reliability, objectivity, and security (NIST, 2021).
- The British Standards Institute (BSI) issued the standard BS 8611 on "Ethics design and application of robots".

It can be seen from the examples listed above that the development of standards on artificial intelligence and ethics are going on with great human intellectual investments by the AI community and the standardization bodies reflecting the supreme importance of the field.

4. Ethical AI software development

4.1. Ethical principles and frameworks

Numerous different organizations, universities and research institutions (e.g., the Turing Institute, Future of Life Institute, Stanford Centre for Internet and Society, Berkman Klein Centre) have composed guidelines, proposals, and overviews on how to develop trustworthy AI systems and applications (Askell, et al., 2019; Future of Life, 2017; Leslie, 2019; Müller, 2020). There are surveys that have analysed the different AI ethical frameworks, approaches and made conclusions on the most important themes and principles, involved in ethical approach of AI (Fjeld et al., 2020; Jobin et al., 2019; Müller, 2020).

One of the first warning on the risks of fast developing AI technologies was issued by the Future of Life Institute with title "Asilomar AI Principles" (Future of Life, 2017). This study contains 23 guidelines for the research and development of artificial intelligence systems. The principles outline AI development issues, ethics and guidelines for the development of beneficial AI and to make beneficial AI development easier. In particular, three categories are discussed: Research (5), Ethics and Values (13) and Longer Term Issues (5). Ethics and values principles are the following: Safety, Failure transparency, Judicial transparency, Responsibility, Value alignment, Human values, Personal privacy, Liberty and privacy, Shared benefit, Shared prosperity, and Human control.

According to Jobin, Ienca and Vayena (2019) a convergence around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy) can be observed. The final conclusion suggests principled, guideline-driven development efforts with substantive ethical analysis and adequate implementation strategies.

A summary and detailed analysis of AI-related ethical themes and principles have been provided by Fjeld et al. based on the policies of 36 countries. Eight main themes (privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility and promotion of human values) and 47 principles have been identified (Fjeld et al., 2020). The authors declare that "AI and robotics have raised fundamental questions about what we should do with these systems, what the systems themselves should do, and what risks they have in the long term. They also challenge the human view of humanity as the intelligent and dominant species on Earth". Note that researchers who investigated the ethical implications of operational research (OR) arrived at a similar typology (Le Menestrel & Van Wassenhove, 2004; Picavet, 2009).

Based on the above analysis it can be stated that fairness, transparency, explainability, responsibility, safety and reliability, security and privacy belong to the most important principles in AI ethics. These ethical principles should be mapped through ethical frameworks into working AI applications.

4.2. Responsible AI approach

Responsibility is a key factor in handling highly automated and autonomous systems. In the relation of AI and ethics, Responsible AI (RAI) represents a methodology that focuses on human responsibility along all phases of the development process. In this context, three approaches can be identified as integrative components of RAI (Dignum, 2018):

- *Ethics by Design* – the technical integration of ethical reasoning capabilities,
- *Ethics in Design* – engineering techniques for evaluating the ethical capabilities of AI systems, and
- *Ethics for Design* – standards and certification processes for providing the integrity of stakeholders during the life-cycle AI systems.

The responsible AI development follows the steps and the theoretical considerations included in ethical frameworks. RAI is a governance framework focusing on designing and implementing ethical, transparent and accountable AI solutions that help maintain individual trust and minimize privacy attack. RAI places humans in the centre and implementing RAI means to match applicable laws, regulations and standards.

A complex definition of RAI is given by Askell, Brundage & Hadfield (2019): "Responsible AI development involves taking steps to ensure that AI systems have an acceptably low risk of harming their users or society and, ideally, to increase their likelihood of being socially beneficial. This involves testing the safety and security of systems during development, evaluating the potential social impact of the systems prior to release, being willing to abandon research projects that fail to meet a high bar of safety, and being willing to delay the release of a system until it has been established that it does not pose a risk to consumers or the public."

When implementing responsible AI there are four main practices (Wang et al. 2020):

- Data governance like transparency, trust building, explain ability.
- Ethically designed solutions to the challenges, like (1) data and cyber security, (2) reducing risk of unethical behaviours and (3) developing ethical mind-set culture for organizations and employees.
- Human centric surveillance/risk control: series of risk control mechanisms in the design, implementation and evaluation phases. Targets include security risks (intrusion-, privacy-, open source-source risks), economic risks (job displacement risk), performance risks (of errors and bias, of black box, and expandability).
- Training and education so that employees and managers can understand the ethical use of AI techniques, including the collection and handling of data.

As RAI is a human-centred methodology, trust plays a vital role in its implementation. A RAI approach can be different with regard to sectors of economy and industry. E.g., the Defence Innovation Unit of Department of Defence (DoD) published a guideline with the title "Responsible AI guidelines in practice" (Dunnmon, Bryce Goodman, Kirechu, Smith & Van Deusen, 2021) for implementing DoD's ethical principles for artificial intelligence for its AI prototype projects.

In AI "black box" and "white box" models can be distinguished. Black box models (as neural networks) may provide great accuracy, but their

inner workings are hard to understand and they rarely give a hint as for the relevance and importance of features used in predictions. White box models are more transparent and can give explanations and interpretations of their outcomes. Explainable AI (XAI) attempts to find explanations for black box models that are too complex to be understood by humans. XAI refers to a set of processes and methods that allow human users to comprehend and trust the results generated by AI algorithms. It helps characterize model accuracy, fairness and transparency as well, so it is a basic method for trust building (Gunning et al., 2019).

5. Key challenges of AI applications in CPPS

5.1. Characteristics of CPPS

Cyber-Physical Production Systems (CPPS) are based on the CPS (Cyber-Physical System) principles: intensive communication and coordination between the physical elements and computational software. According to Monostori et al. (2016) "CPPS consists of autonomous and cooperative elements and sub-systems that are connected based on the context within and across all levels of production, from processes, through machines and up to production and logistics networks". The main characteristics of CPPS are their (1) computational intelligence, (2) continuous communication, and (3) resilience to the changing and uncertain production environment through flexible control.

In case of conventional production systems adaptive control is a proper approach to control both deterministic and continuous-time systems and stochastic discrete-time systems as well (Anuradha, Annaswamy & Fradkov, 2021), but CPPSs need novel control approaches. Even though AI technologies can contribute to these control systems in many ways, the involvement of humans in CPPS control remains a focal point, especially if human-machine cooperation and collaboration is to be facilitated. The phases of evolution with increasing involvement of AI technologies along with human participation in the control of manufacturing system are shown in Fig. 3, extending the description in Cardin, (2019). Similar levels are defined by the IEEE Standard for Transparency of Autonomous Systems (IEEE, 2022).

- **Level 1**. *Control by Human*: Here human has full, direct control over the system which provides only data to the user who is in charge of making and implementing all the decisions. Traditional systems like small-scale job shops, conventional milling machines with manual functions, or manual material handling systems are operated with this elementary level of control.
- **Level 2**. *AI Assisted Control*: The system has autonomy in limited fields, control decisions are made by the human who is in charge of most of the decisions. E.g., a CNC machine can operate with a pre-programmed technology following a schedule which was developed by an AI system. Also, robots can perform part handling tasks with limited functions based on AI path planning.
- **Level 3**. *AI Advisory Control*: The system guides the human during its task execution by taking most of the decisions and leaving the functions of adaptation to the human. Such function is predictive maintenance when from IIoT data the system suggests repair date, time, object(s) and action(s) of repair, but leaves final decisions at the workmen.
- **Level 4**. *AI Collaborative Control*: A partnerships or teamwork where leadership and responsibility are shared by the partners dynamically, depending on the actual tasks under execution. In a collaboration relationship the partners have common goals and help each other to do the right thing. In production, collaborative robots (cobots) work with humans in some way – either as an assistant, or as a leader (Wang et al., 2019). Collaborative robots are designed and programmed to respond to human instructions and actions in a shared workspace, warranting human's safety (Horváth & Erdős, 2017).
- **Level 5**. *AI Active, Independent Control*: Autonomous Control Systems (ACS) use a combination of model-based engineering, AI technologies (typically, machine learning), and data acquisition to enable self-governed, autonomous control functions with little to no human intervention for extended periods in an uncertain or contested environment. The human has only supervision role in the CPPS, which can make all the necessary decisions without any human intervention. A fleet of autonomous mobile robots providing internal logistics services for production is a good example for such advanced control (Beregi, Pedone, Háy & Váncza, 2021).

Of course, most challenging is Level 5 control, where a CPPS is typically a highly distributed, labyrinthine-like collaborative network. When balancing between autonomy and cooperation, conflicts have to be resolved time and again, in a flexible and efficient way. The assignment of responsibilities to different agents in the system can help finding a resolution (Karnouskos, Ribeiro, Leitão, Lüder & Vogel-Heuser, 2019). Furthermore, CPPS control calls for special faculties like self-configuration, self-adjustment, and self-optimization (see Fig. 4). Advanced AI – multi-agent systems in particular – can provide key enabling technologies for all of that (Monostori et al., 2006), but these extensions demand new security, safety and trust models. The ethics in CPPS boils down to the demand for trustworthy systems, where trust between humans and machines should be mutual (Radanliev, De Roure, Van Kleek, Santos & Ani, 2020; EC, 2020).

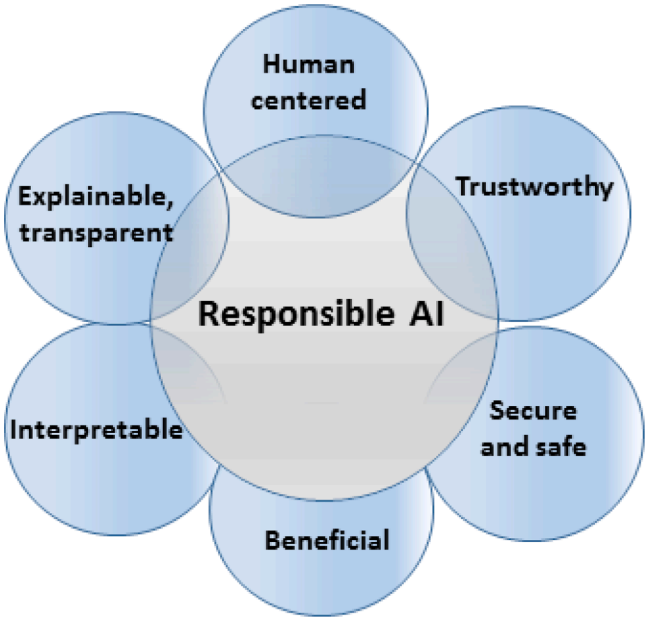

Fig. 2. The components of Responsible AI.

5.2. Challenges in CPPS

There are numerous challenges in production due to the introduction of CPPS but in the followings the focus is set only to those which can be connected to ethical aspects (Scientific Foresight Unit, 2016). The main challenges are originating from the function, structure and operation mood of CPPS, namely:

- Communication, moving and storage of big amount of data – security challenge.
- Generating, collecting and handling data – privacy challenge.
- Close and permanent cooperation, collaboration of different agents (machines, computers and humans) results safety problems (physical as well).
- Control and predictions on processes, machines – trust in algorithms.

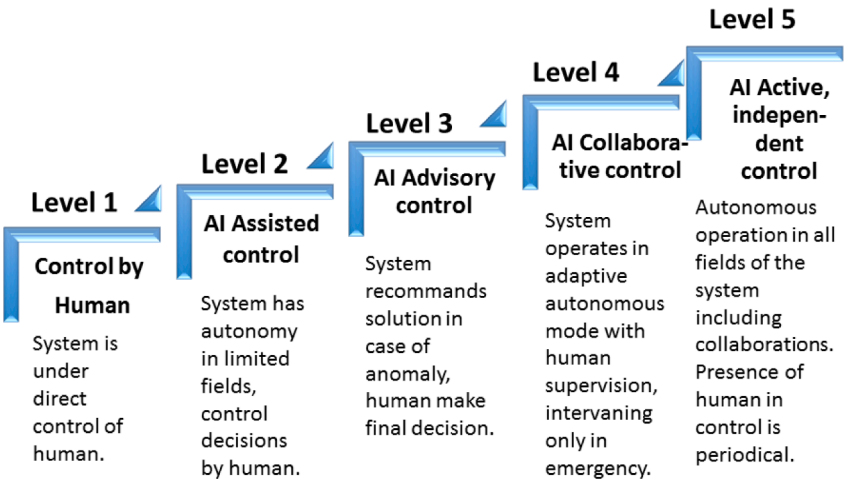

Fig. 3. Evolution of inclusion of AI in control systems.

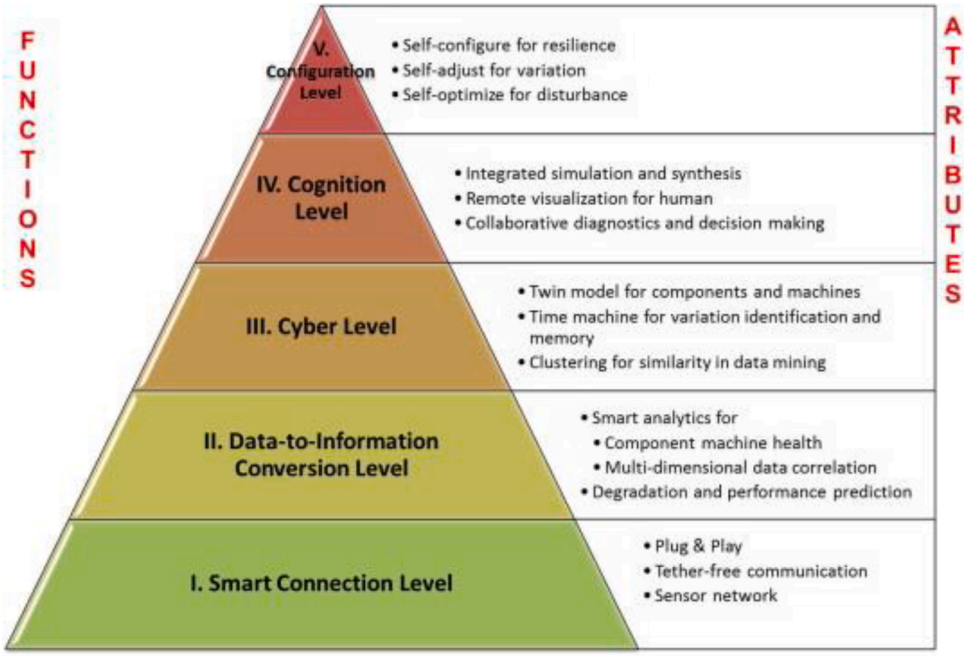

Fig. 4. 5C architecture for implementation of cyber-physical system (Lee et al., 2015).

The listed challenging components are essential in building trust between humans, machines and computer systems, so they are building blocks of CPPS ethics when implementing trustworthy systems. In case of building ethical systems using the control systems, especially AI-based ones, it is essential to take into consideration how security, privacy and safety can be build.

A detailed analysis of the challenges of cyber-physical manufacturing enterprises of the future have been done by Panetto et al. (2019). Four Grand Challenges have been defined and each have been decomposed to further categories; Business, Knowledge, Applications, Communications (ICT).

- Grand Challenge 1. CPPS-based manufacturing plant control
- Grand Challenge 2. Resilient digital manufacturing networks, collaborative control for Industry 4.0 and cyber-physical supply chains
- Grand Challenge 3. Cyber-physical system-of-systems interoperability
- Grand Challenge 4. Interdependent networked systems and data analytics for decision support

AI application opportunities and challenges have been identified explicitly in three cases: (1) AI and data-driven business, (2) symbolic artificial intelligence and software agents, and (3) cobots and new human-machine interaction with robots. However, numerous other AI applications can be involved in the different categories of challenges, like Challenge 1.1. which is about CPPS-based autonomous shop-floor systems, or Challenge 4.2 on AI and data-driven modelling in multi-level, multi-scale systems.

5.3. *Responsible AI system development in CPPS*

Responsible AI is a methodology for translating ethical principles (e.g., algorithmic fairness) into practical, measurable metrics for industrial AI applications. As summarized in Fjeld, Nele, Hannah, Nagy and Srikumar, (2020), fairness, transparency, explainability, responsibility,

safety and reliability, security and privacy belong to its core notions. In case of practical applications, the ranking of these principles should be the first step of the development process. The RAI process is usually divided into four phases: planning, development, deployment, and tracing the operation of the system. Tracing involves the continuous collection of results, tested by developers, as well as the recording of and reflections to customer feedback.

Instantiations of the RAI methodology are the different frameworks and toolkits (Khargonekar & Sampath, 2020). These can be advanced development environments or traditional guidelines. The principles have to be adapted after analysing the actual organizational, technical, operational and legal requirements and expectations, since companies have typically different policies with respect to artificial intelligence. An essential step towards responsible AI is the clear-cut assignment of responsibilities: who is liable if something goes wrong during the operation of the application.

Dominant number of AI applications are operational already in the financial sector (fintech) where data is available in great quantity and good quality. The broad usage of complex AI applications in machine industry started only later, mostly because of data collection and data quality problems. The wide introduction of Industrial Internet of Things (IIoT) and advanced sensor and networking technologies can now provide sufficient data in real time, so the number of industrial AI applications started to increase too.

However, the AI technologies applicable in CPPS need different RAI approaches, depending on the ultimate goals of their use. E.g., in case of object recognition, it makes a big difference whether human beings or objects of production are to be identified (see also Table 3). The diverse requirements should be specified already in the first phase of system design. Table 3 gives some examples for various application possibilities of AI in CPPS taken from our recent practice. The application of RAI method for CPPS needs in all these cases a very thorough analysis of the demands, the explicit expression and prioritization of principles and the selection of the proper software tools and toolkits.

The references introduced in Table 3, are different AI-based applications, mostly deployed in our pilot CPPS. The system contains a variety of elements, like single robot arms, production assembly lines, Autonomous Mobile Robots (AMR) fleet (with and without cobot on), collaborative robots and human operated components, such as the warehouse and also a digital work assistance system. Various AI technologies were employed for vision-based situation recognition, machine learning, data analysis, gesture-based robot control, autonomous vehicle control, and automated robot programming. At the time when the pilot systems were established, AI standards were just getting conceived, so they could not be taken into consideration. The existing non-direct AI standards were applied and validated in several use-cases for communication interoperability in the Manufacturing Execution Systems (MES) architecture that connects the components of the system (Beregi et al., 2021) as well as the security standard GDPR. The two major standardization frameworks for industrial Internet architectures were also investigated: the Industrial Internet Reference Architecture (IIRA) and the Reference Architectural Model Industry (RAMI 4.0) (Pedone & Mezgár, 2018). Now a compliance check is on its way to investigate how these control systems match the actual AI standards in system design, and the transparency requirements of robot control software (IEEE, 2021). The predefined scenarios together with the transparency score sheets of (IEEE, 2022) facilitate effectively the checking of safety demands.

As we see now, applying IEEE 7000–2021 and IEEE 7000–2022 can provide a good guidance during system design and transparency analysis when including also ethical AI considerations. Advantages are that the system will be reliable, no aspects will be missed, and the subsystems will match seamlessly, satisfying the ethical demands both on equipment and on system level while reducing development time as well.

6. Discussion

Based on the content of the previous section in what follows we summarize the key generic and specific dilemmas and also our recommendations related to the realization of ethical behaviour in Cyber-Physical Production Systems that apply some AI technologies.

The main generic dilemmas of the ethical engineering applications of AI are the following:

- The emergent behaviour of a CPPS which consists of autonomous agents implies real risks because its globally correct operation can hardly be warranted. Only some uncertain predictions, typically on the basis of simulation results, can be given.
- The acceptance of AI technology by the users depends on their trusting in these systems that is influenced in a high degree both by the related legal environment, standards and the technical background. People can hardly be convinced to trust in a “black box” technology. Hence, features like safety, security, transparency, explainability are of crucial importance.
- In case of real autonomous systems, dependability/reliability of the system and its ability to handle potential malfunctions and failures, i.e., features like failure-free and reliable operation, self-monitoring and repair are keys for social acceptance. Providing these faculties in an autonomous system can be a big challenge.
- Given limited computational resources, how can all these properties be warranted?

Table 3
Possible application fields of AI in CPPS.

	AI technologies								
Application field	Worker recogn.	Work-piece recogn.	Machine recogn.	Workp. + AGV recogn.	Prediction	Control	Data analysis	Planning	Collaboration
Assembly	Tsutsumi et al., 2020	Tipary & Erdős, 2021				(Beregi et al., 2019)		Kardos, Kovács & Váncza, 2020	Wang et al., 2019
AGV operation		Beregi et al., 2021	Beregi et al., 2021	Beregi et al., 2021		Beregi et al., 2021	(Gyulai, Bergmann & Váncza, 2020)	Bergmann, Gyulai & Váncza, 2021	
Scheduling					Frye, Gyulai, Bergmann & Schmitt, 2019	(Gyulai, Pfeiffer & Bergmann, 2020)			
Worker monitoring	Tsutsumi et al., 2020								
Robotics						(Horváth and Erdős, 2017)		Erdős, Kovács & Váncza, 2016	Kemény, Váncza, Wang, & Wang, 2021

- In a hybrid autonomous system where human and machine agents are working collaboratively, how can collective, system-level responsibilities be assigned to individuals?

Conflicting requirements expose some dilemmas as for defining ethical standards to AI systems in a CPPS application:

- CPPS changes the conventional characteristics of production, where (mass) customization, responsiveness, flexibility and robustness became main features. The production of unique, customer designed products and services will be more and more dominant. Should CPPS be legally bound to some ethical principles in the customization and individualization of products and attached services?
- Privacy versus efficiency on the shop-floor exposes a specific dilemma which has ethical repercussions as well. Is it ethical to supervise workers with cameras, analysing their behaviour, giving remote commands them on what, where, and how to do? Note that in case of perceived fatigue the actual workload could also be alleviated. How can conflicting aspects of ethics, efficiency and business be resolved?
- How to assign responsibility for system errors, how to share collective and individual responsibilities? In case of an accident in a factory originating from the malfunction of a smart machine who should the blame be assigned: the worker or the owner or the machine manufacturer or its software developer or the learning data provider (supposing a learning model) or the algorithm developer or who else?
- How safety standards and certification of products can be considered in the context of increased customisation? If a product doesn't comply with safety standards, who is liable, the manufacturer or the designer?
- Privacy will become a more and more critical issue with the proliferation of CPPS. Large amounts of data will be collected at each stage of the manufacturing process from all entities. How data protection will be solved and what data should companies be allowed to capture on customers and employees?
- On the high level, what is the role and responsibility of policy makers, politicians, legal system administrations, standard organizations in setting up a technologically, economically and ethically feasible framework for autonomous CPPS systems?

For the time being, on the basis of all the above considerations we can suggest some recommendations for the uptake of ethical AI applications in CPPS.

- The responsible AI (RAI) human-centred approach provides an implementation of ethical frameworks and supports the life-cycle development of trustworthy, ethical AI applications that can be applied in a CPPS environment as well ("ethics by design"). Even though the RAI approach can be set up differently in various sectors of the economy or at different companies, all need a reference RAI process model.
- Development of standards related to AI and ethics are going on with great intensity, reflecting the importance of this field. However, research in ethical AI is still in infancy, these kinds of investigations have to be accelerated. However, as it seems now, contemporary theoretical discourse on ethics can motivate even new operational principles.
- In case of autonomous systems special tests have to be applied at every phase of design and development so as to check the reliability of the systems and their elements against excessive risks. Such stratified testing should precede the start of a real-life implementation and employment.
- AI strategies developed by many countries – which include proposals for new AI related laws and development strategies as well – should be unified and generic and be elaborated for their application in CPPS. However, these frameworks should be updated regularly according to the progress of AI and production technologies.

7. Conclusions

The evolution of artificial intelligent technologies has extremely accelerated in the last decade and contributed significantly also to the development of Cyber-Physical Production Systems. In our view, this turn was initiated and driven mostly by the emerging task-oriented agent-based model of AI. However, AI can transform the relations between humans, devices and society in an undefined, so far mostly unknown way and extent, generating important ethical, legal and standardisation issues. The real risks and their mitigations are not exactly known, as there are mostly only uncertain predictions on how autonomous AI-based systems will behave and interact.

The acceptance of AI technologies by the users depends on their trusting in these systems. This trust is determined in a high degree both by the related legal environment and standards and by the technical background which can warrant critical features like safety, security, transparency, and explainability. The paper outlined how user's trust can be generated and maintained by technical solutions, and also explained the important role of laws and standards in this process. The user's trust in a system is a basic demand especially in industrial AI applications where both the physical safety of humans and the consistent operation of autonomous agents have to be warranted.

The paper gave a structured overview on ethics related issues and the status of AI technology standardisation that define the broad usability of industrial AI systems. An overview of AI applications in CPPS was presented as well, introducing the Responsible AI approach for developing ethical AI systems. In the last section dilemmas and recommendations were summarized with a special respect to CPPS. The real threat of AI is not its taking the control over the world but the undetected errors, failures and misuses resulting from wrong and unexplained decisions.

In conclusion, we can but suggest that the ethical aspects and principles should be taken into consideration from the outset of system development, taking the "ethics by design" approach. Otherwise, improperly developed AI systems would cause (physical) damages both in humans, machines and in the environment and additionally a loss of trust in the existing and emerging AI technologies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been supported by the European H2020 EPIC grant no. 739592, and the Hungarian NRDIO grant no. TKP2021-NKTA-01.

References

Anscombe, G. E. (1969). Modern moral philosophy. *The is-ought question* (pp. 175–195). London: Palgrave Macmillan.

Anuradha, M., Annaswamy, A., & Fradkov, L. (2021). A historical perspective of adaptive control and learning. *Annual Reviews in Control, 52*, 18–41.

Askell, A., Brundage, M., & Hadfield, G. (2019). The role of cooperation in responsible AI development. In: ArXiv (July 2019). arXiv: 1907.04534. http://arxiv.org/abs/1907.04534.

Beregi, R., Pedone, G., Háy, B., & Váncza, J. (2021). Manufacturing execution system integration through the standardization of a common service model for cyber-physical production systems. *Applied Sciences, 11*, 7581.

Beregi, R., Pedone, G., & Mezgár, I. (2019). A novel fluid architecture for cyber-physical production systems. *International Journal of Computer Integrated Manufacturing, 32* (4–5), 340–351.

Bergmann, J., Gyulai, D., & Váncza, J. (2021). Adaptive AGV fleet management in a dynamically changing production environment. *Procedia Manufacturing, 54*, 148–153.

BEUC. (2020). Artificial intelligence: What consumers say: *Findings and policy recommendations of a multi-country survey on AI. BEUC*. Report, June 2020, p. 11.
Bosch, (2020). *In brief: Bosch code of ethics for AI*. February 19, 2020 PI 11094 RB Cwi/BT.
Brandt, R. (1990). The science of man and wide reflective equilibrium. *Ethics, 100*, 259–278.
Byrne, G., Damm, O., Monostori, L., Teti, R., van Houten, F., Wegener, K., & Sammler, F. (2021). Towards high performance living manufacturing systems – A new convergence between biology and engineering. *CIRP Journal of Manufacturing Science and Technology, 34*, 6–21.
Byrne, G., Dimitrov, D., Monostori, L., Teti, R., & van Houten, F. (2018). Biologicalisation: Biological transformation in manufacturing. *CIRP Journal of Manufacturing Science and Technology, 21*, 1–32.
Cambridge Dictionary, (2021). https://dictionary.cambridge.org/dictionary/english/.
Cardin, O. (2019). Classification of cyber-physical production systems applications: Proposition of an analysis framework. *Computers in Industry, 104*, 11–21.
China Artificial Intelligence Standardization (2020). White Paper. Translation, May 12, 2020, https://cset.georgetown.edu/research/artificial-intelligence-standardization-white-paper/.
Cihon, P. (2019). Standards for AI governance: International standards to enable global coordination in AI research & development. Technical Report, University of Oxford.
Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*, 1–3.
Ding, J. (2018). *Deciphering China's aI dream – the context, components, capabilities, and consequences of china's strategy to lead the world in AI. Centre for the governance of AI*. Oxford, UK: Future of Humanity Institute, University of Oxford. March 2018.
Dunnmon, J., Bryce Goodman, B., Kirechu, P., Smith, C., & Van Deusen, A. (2021). *Responsible AI guidelines in practice*. Defense innovation unit, DoD, https://www.diu.mil/responsible-ai-guidelines.
EC, European Commission (2018a). *Communication artificial intelligence for Europe*.
EC, European Commission (2018b). *Ethic guidelines for trustworthy AI*, High-Level Expert Group on Artificial Intelligence (AI HLEG).
EC, European Commission. (2019). *A definition of AI: Main capabilities and scientific disciplines*. Brussels: High-Level Expert Group on Artificial Intelligence.
EC, European Commission (2020). *Report on the safety and liability implications of artificial intelligence, the internet of things and robotics*.
EC, European Commission (2021a). *Proposal for a regulation of the European parliament and of the council laying harmonised rules on artificial intelligence (AI Act) and amending certain legislative acts*. Document 52021PC0206.
EC, European Commission (2021b). *Proposal for a regulation of the European parliament and of the council on machinery products*. Document 52021PC0202.
EDN. Basics of the IEEE Standardization process. https://www.edn.com/basics-of-the-ieee-standardization-process/.
Erdős, G., Kovács, A., & Váncza, J. (2016). Optimized joint motion planning for redundant industrial robots. *CIRP Annals – Manufacturing Technology, 65*(1), 451–454.
Fisher, M., List, C., Slavkovik, M., & Winfield, A. (2016). Engineering moral machines. *Informatik-Spektrum, 39*, 467–489.
Fjeld, J., Nele, A., Hannah, H., and Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. (January 15, 2020, Berkman Klein Center Research Publication No. 2020-1.
Frye, M., Gyulai, D., Bergmann, J., & Schmitt, R. H. (2019). Adaptive scheduling through machine learning-based process parameter prediction. *MM Science Journal*, (Special Issue on HSM2019), 3060–3066.
Future of Life, (2017). *Asilomar AI principles*. https://futureoflife.org/ai-principles/?cn-reloaded=1.
Gibson, Dunn (2022). *2021 artificial intelligence and automated systems annual legal review*. January 20, 2022, https://www.gibsondunn.com/2021-artificial-intelligence-and-automated-systems-annual-legal-review/.
Gillespie, N., Lockey, S., & Curtis, C. (2021). *Trust in artificial intelligence: A five country study*. The University of Queensland and KPMG Australia. doi: 10.14264/e34bfa3.
Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI-explainable artificial intelligence. *Science Robotics, 4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120
Gyulai, D., Bergmann, J., & Váncza, J. (2020a). Adaptive network analytics for managing complex shop-floor logistics systems. *CIRP Annals – Manufacturing Technology, 69*(1), 393–396.
Gyulai, D., Pfeiffer, A., & Bergmann, J. (2020b). Analysis of asset location data to support decisions in production management and control. *Procedia CIRP, 88*, 197–202.
Hockfield, S. (2019). *The age of living machines: How biology will build the next technology revolution*. New York, London: W. W. Norton & Company.
Hooker, J., & Kim, T. W. (2019). Truly autonomous machines are ethical. *AI Magazine, 40*(4), 66–73.
Horváth, G., & Erdős, G. (2017). Gesture control of cyber-physical systems. *Procedia CIRP, 63*, 184–188.
IEEE. (2021). IEEE standard model process for addressing ethical concerns during system design. *IEEE*, 7000-2021, p. 80.
IEEE, (2022). *IEEE standard for transparency of autonomous systems*. IEEE Std 7001™-2021.
Intelligence Community, (2020). *Artificial intelligence ethics framework for the intelligence community*. INTEL.gov Version 1.0, June 2020, https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community.
Ipsos, (2022). *Global opinions and expectations about AI*. January 2022.
ISO, (2020). *Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*. ISO/IEC TR 24028:2020, https://www.iso.org/standard/77608.html.
ISO, (2021). *Artificial intelligence*. ISO/IEC JTC 1/SC 42, https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399.
Kahneman, D. (2011). *Thinking, fast and slow*, Farrar, Straus and Giroux.
Kardos, C., Kovács, A., & Váncza, J. (2020). A constraint model for assembly planning. *Journal of Manufacturing Systems, 54*, 196–203.
Karnouskos, S., Ribeiro, L., Leitão, P., Lüder, A., & Vogel-Heuser, B. (2019). Key directions for industrial agent based cyber-physical production systems. In *2019 IEEE international conference on industrial cyber physical systems (ICPS), 2019* (pp. 17–22).
Kemény, Zs., Váncza, J., Wang, L., & Wang, X. V. (2021). Human–Robot collaboration in manufacturing: A multi-agent view. In L. Wang, X. V. Wang, J. Váncza, & Zs. Kemény (Eds.), *Advanced human-robot collaboration in manufacturing* (pp. 3–41). Cham, Switzerland: Springer.
Keng, S., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74–86.
Khargonekar, P. P., & Sampath, M. (2020). A framework for ethics in cyber-physical-human systems. *IFAC-PapersOnLine, 53*(2), 17008–17015.
Kuipers, B. (2020). Perspectives on ethics of AI: Computer science. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI*. Oxford, UK: Oxford University Press.
Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*, Viking.
Le Menestrel, M., & Van Wassenhove, L. N. (2004). Ethics outside, within, or beyond OR models? *European Journal of Operational Research, 153*(2), 477–484.
Lee, J., Bagheri, B., & Kao, H.-. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*, 18–23.
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*, 50–80.
Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529.
McFadden, M., Jones, K., Taylor, E., & Osborn, G. (2021). *Harmonising artificial intelligence: The role of standards in the EC AI regulation*. Oxford Information Labs, Working paper 2021.5, https://oxcaigg.oii.ox.ac.uk.
Luhmann, N. (1979). *Trust and power*. New York: John Wiley & Sons.
Mezgár, I. (2021). From ethics to standards: An overview of AI ethics in CPPS. *IFAC-PapersOnLine, 54*(1), 723–728.
Ministry of Justice, (2016). *The Swedish law-making Process*. https://www.government.se/49c837/contentassets/4490fe7afcb040b0822840fa460dd858/the-swedish-law-making-process.
Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhard, G., & Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals – Manufacturing Technology, 65*(2), 621–641.
Monostori, L., & Váncza, J. (2020). Towards living manufacturing systems. *Procedia CIRP, 93*, 323–328.
Monostori, L., Váncza, J., & Kumara, S. R. (2006). Agent-based systems for manufacturing. *CIRP Annals – Manufacturing Technology, 55*(2), 697–720.
Monroe, D. (2014). Neuromorphic computing gets ready for the (really) big time. *Communications of the ACM, 57*(6), 13–15.
Müller, V. C., & Zalta, E. (2020). Ethics of artificial intelligence and robotics. *The Stanford encyclopedia of philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University (Winter 2020 Edition) https://plato.stanford.edu/archives/win2020/entries/ethics-ai.
Nagel, T. (2021). Types of intuition – Intimations of morality. *London Review of Books, 43*(11), 1–11.
National Governance Committee of China. (2021). *Ethical norms for the new generation artificial intelligence*. https://ai-ethics-and-governance.institute/2021/09/27/the-ethical-norms-for-the-new-generation-artificial-intelligence-china/.
Neudert, L-M., Knuutila, A., & Howard, P.N. (.2020). *Global attitudes towards AI, machine learning & automated decision making*. Working paper 2020.10, Oxford Commission on AI & Good Governance, 10 pp.
Neugebauer, R., Ihlenfeldt, S., Schliesman, U., Hellmich, A., & Noack, M. (2019). A new generation of production with cyber-physical systems: Enabling the biological transformation in manufacturing. *Journal of Machine Engineering, 19*(1), 5–15.
NIST, (2021). *AI standards*. https://www.nist.gov/topics/artificial-intelligence/ai-standards.
Nocetti, J. (2020). *The outsider: Russia in the race for artificial intelligence*. Russie.Nei. Reports, No.34. Ifri, December 2020.
OECD. (2019). AI policies and initiatives. *Artificial intelligence in society*. Paris: OECD Publishing. https://doi.org/10.1787/cf3f3be0-en
OECD, (2021). *Database of national AI policies, Powered by EC/OECD*. https://oecd.ai.
Panetta, K. (2018). 5 trends emerge in the gartner hype cycle for emerging technologies. https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/.
Panetto, H., Iung, B., Ivanov, D., Weichhart, G., & Wang, X. (2019). Challenges for the cyber-physical manufacturing enterprises of the future. *Annual Reviews in Control, 47*, 200–213.
Pedone, G., & Mezgár, I. (2018). Model similarity evidence and interoperability affinity in cloud-ready. Industry 4.0 technologies. *Computers in Industry, 100*, 278–286
Picavet, E. (2009). Opportunities and pitfalls for ethical analysis in operations research and the management sciences. *Omega, 37*(6), 1121–1131.
Pope Francis, (2020). *The Pope's monthly intentions for 2020*, November, https://www.usccb.org/prayer-and-worship/prayers-and-devotions/the-popes-monthly-intention.
Radanliev, P., De Roure, D., Van Kleek, M., Santos, O., & Ani, U. (2020). Artificial intelligence in cyber physical systems. *AI & Society, 36*(3), 783–796.

Rothenberger, L., Fabian, B., & Arunov, E. (2019). Relevance of ethical guidelines for artificial intelligence – a survey and evaluation. In *Proc. of the 27th European conference on information systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019*. https://aisel.aisnet.org/ecis2019_rip/26.

Rousseau, D. M., Sitkin, S. B., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-disciplinary view of trust. *Academy of Management Review, 23*(3), 393–404.

Russell, S. (2016). Should we fear super smart robots? *Scientific American, 314*, 58–59.

Russell, S., & Wefald, E. (1991). *Do the right thing – studies in limited rationality*. Cambridge, MA: MIT Press.

Scheutz, M. (2017). The case for explicit ethical agents. *AI Magazine, 38*(4), 57–64.

Schmelzer, R. (2020). Worldwide AI laws and regulations 2020. *Cognilytica Research*. Document ID: CGR-REG20, p. 97.

Scientific Foresight Unit, (2016). *Ethical aspects of cyber-physical systems. Scientific foresight study*. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/563501/EPRS_STU%282016%29563501_EN.pdf.

Sifakis, J. (2019). Can we trust autonomous systems? Boundaries and risks. InY. F. Chen, C. H., Ch.eng, & J., Esparza. (eds), *Automated technology for verification and analysis, lecture notes in computer science*, vol. 11781, pp. 65–78, Springer.

Svegliato, J., Nashed, S., & Zilberstein, S. (2020). An integrated approach to moral autonomous systems. In *Proc. of the 24th European conference on artificial intelligence (ECAI 2020), IOS Press* (pp. 2941–2942).

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence, 2*(1), 10–12.

Tipary, B., & Erdős, G. (2021). Generic development methodology for flexible robotic pick-and-place workcells based on digital twin. *Robotics and Computer-Integrated Manufacturing, 71*(9), Article 102140.

Trentesaux, D., & Karnouskos, S. (2022). Engineering ethical behaviors in autonomous industrial cyber-physical human systems. *Cognition, Technology & Work, 24*, 113–126.

Tsutsumi, D., Gyulai, D., Takács, E., Bergmann, J., Nonaka, Y., & Fujita, K. (2020). Personalized work instruction system for revitalizing human-machine interaction. *Procedia CIRP, 93*, 1145–1150.

US Government Documents Collection, (2022). *The lawmaking process: How a bill becomes a law*. https://guides.nyu.edu/govdocs/lawmaking.

US National Science and Technology Council, (2016). *National artificial intelligence research and* development strategic plan, October 2016.

Veale, M., & Borgesius, Z. (2021). Demystifying the draft EU artificial intelligence act (July 31, 2021) *Computer Law Review International, 22*(4), 97–112.

Wallach, W., & Vallor, S. (2020). Moral machines. In S. M. Liao (Ed.), *Ethics of artificial intelligence*. Oxford, UK: Oxford University Press.

Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., et al. (2019). Symbiotic human-robot collaborative assembly. *CIRP Annals – Manufacturing Technology, 68*(2), 701–726.

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an understanding of responsible artificial intelligence practices. In T. X. Bui (Ed.), *Proceedings of the 53rd Hawaii international conference on system sciences. Hawaii international conference on system sciences (HICSS 2020)* (pp. 4962–4971).

Xu, J., Le, Kim, Deitermann, A., & Montague, E. (2014). How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology. *Applied Ergonomics, 45*(6), 1495–1503.

Yeung, K. (2020). Recommendation of the Council on Artificial Intelligence (OECD). *International Legal Materials, 59*(1), 27–34. https://doi.org/10.1017/ilm.2020.5

Zachariadis, I. (2019). Standards and the digitalisation of EU industry economic implications and policy developments. *European Parliamentary Research Service*, 1–8.

Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., et al. (2020). The combination of brain-computer interfaces and artificial intelligence: Applications and challenges. *Annals of Translational Medicine, 8*(11), 712.