Journal of Machine Engineering, 2020, Vol. 20, No. 4, 161–184
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received: *08 October 2020* / Accepted: *28 November 2020* / Published online: *04 December 2020*

human role, artificial intelligence,
machine learning, manufacturing singularity,
intelligent machine architecture,
cyber-physical systems, industry 4.0

Goran D. PUTNIK[1,3*]
Vaibhav SHAH[2,3]
Zlata PUTNIK[3]
Luis FERREIRA[4]

MACHINE LEARNING IN CYBER-PHYSICAL SYSTEMS AND *MANUFACTURING SINGULARITY* – IT DOES NOT MEAN TOTAL AUTOMATION, HUMAN IS STILL IN THE CENTRE:

Part I – *MANUFACTURING SINGULARITY* AND AN INTELLIGENT MACHINE ARCHITECTURE

In many popular, as well scientific, discourses it is suggested that the "massive" use of Artificial Intelligence, including Machine Learning, and reaching the point of "singularity" through so-called Artificial General Intelligence (AGI), and Artificial Super-Intelligence (ASI), will completely exclude humans from decision making, resulting in total dominance of machines over human race. Speaking in terms of manufacturing systems, it would mean that there will be achieved intelligent and *total* automation (once the humans will be excluded). The hypothesis presented in this paper is that there is a limit of AI/ML autonomy capacity, and more concretely, that the ML algorithms will be not able to became totally autonomous and, consequently, that the human role will be indispensable. In the context of the question, the authors of this paper introduce the notion of the *manufacturing singularity* and an *intelligent machine architecture* towards the manufacturing singularity, arguing that the intelligent machine will be always human dependent, and that, concerning the manufacturing, the human will remain in the centre of Cyber-Physical Systems (CPS) and in I4.0. The methodology to support this argument is inductive, similarly to the methodology applied in a number of texts found in literature, and based on computational requirements of inductive inference based machine learning. The argumentation is supported by several experiments that demonstrate the role of human within the process of machine learning. Based on the exposed considerations, a generic architecture of intelligent CPS, with embedded ML functional modules in multiple learning loops, in order to evaluate way of use of ML functionality in the context of CPPS/CPS. Similarly to other papers found in literature, due to the (informal) inductive methodology applied, considering that this methodology doesn't provide an absolute proof in favour of, or against, the hypothesis defined, the paper represents a kind of position paper. The paper is divided into two parts. In the first part a review of argumentation from literature, both in favor of and against the thesis on the human role in future, is presented. In this part a concept of the *manufacturing singularity* is introduced, as well as an *intelligent machine*

[1] University of Minho, School of Engineering, Department of Production and Systems Engineering, Portugal,
[2] University of Minho, School of Engineering, Department of Information Systems, Portugal
[3] ALGORITMI Research Centre, Universidade do Minho, Portugal
[4] Polytechnic Institute of Cávado and Ave, School of Technology, Portugal
[*] E-mail: putnikgd@dps.uminho.pt
https://doi.org/10.36897/jme/131000

architecture towards the manufacturing singularity is presented, arguing that the intelligent machine will always be human dependent, and that, concerning the manufacturing, the human will remain in the centre. The argumentation is based on the phenomenon related to computational machine learning paradigm, as intrinsic feature of the AI/ML, through the inductive inference based machine learning algorithms, whose effectiveness is conditioned by the human participation. In the second part, an architecture of the Cyber-Physical (Production) Systems with multiple learning loops is presented, together with a set of experiments demonstrating the indispensable human role. Also, a discussion of the problem from the manufacturing community point of view on future of human role in Industry 4.0 as the environment for advanced AI/ML applications is included in this part.

1. INTRODUCTION[1]

One of the main requirements for designing and operating new manufacturing devices and systems within the concept of Industry 4.0 (I4.0) is to embed Artificial Intelligence and Machine Learning (AI/ML) functionalities in virtually any single component, making, in fact, I4.0 based systems the systems composed almost exclusively from so called "smart objects"[2]. Such requirement, of the "massive" use of AI/ML, naturally rises a number of questions, of which one of the most important is: where are the limits?

Different types of limits could be defined, depending of the context of considerations, but, phenomenologically, the most important question, from which all other questions stem, is the question if there is the limit of AI/ML in comparison with the human intelligence. The consequence of non-existence of this limit of AI/ML is that the machines could become autonomous, and in the limit, *totally* autonomous, up to the exclusion of humans from decision making from any single issue, which finally could result, by some prominent personalities, even in extinction of human race.

However, there is no positive response to this question. Considering literature, there could be found totally opposite positions regarding this most important question: where are the limits? Or in other words, if there is a limit of AI/ML in comparison with the human intelligence. All positions found, regarding this truly big question, are based on inductive argumentation[3]. As the inductive methodology, by some philosophy of science models, doesn't guarantee positive conclusions, all conclusions based on induction virtually could be considered only as positions or as guidance to orient further research and applications. This fact could actually explain co-existence of the opposite positions regarding the question of AI/ML capacity as compared with human capacity.

From manufacturing engineering point of view, the question has interest to be considered in the context of designing and operation of future manufacturing devices (machines) and systems, especially within the context of newly promoted I4.0. Concerning the I4.0, and in particular Cyber-Physical Production Systems (CPPS, or, further, CPS) as one of the most important I4.0 constructs (also, models or instruments), the question is if the AI/ML could learn, and subsequently generate and operate the (intelligent) control

[1] The paper is based on the Keynote Lecture presented on the 31st CIRP Sponsored Conference on *Supervising and Diagnostics of Machining Systems*, 08-12 March 2020, Karpacz, Poland.

[2] It should not be understood that the use and embedding of AI/ML in industrial devices is the only determinant of I4.0. There are other features that, together with AI/ML, as well as the way of their relationship, determine I4.0, and make I4.0, but these will be not consider in this paper as their consideration doesn't affect the main hypothesis of this paper.

[3] Although inductive reasoning, i.e. 'inductive inference', is one of the ´classical' inference methodologies, and a regular scientific methodology, its value is questioned by some philosophy of science models, e.g. by Popper´s model.

programs, without human role within the process, i.e. without human intervention[4]. If the answer would be 'yes', it would mean a strong suggestion of possible autonomy for more and the most complex tasks, e.g. for autonomous decisions on production at all. This scenario would totally exclude human from the production, and ultimately from running economy. This scenario would be just a part, in the domain of production, of the most extreme vision of the AI/ML capacity which would result in extinction of human race (see below in the second chapter).

The hypothesis presented in this paper is that there is a limit of AI/ML autonomy capacity, and more concretely, that the ML algorithms will be not able to become totally autonomous and, consequently, that the human role will be indispensable. In the context of this question, the authors of this paper introduce the notion of *manufacturing singularity* and an *intelligent machine architecture* towards the manufacturing singularity, arguing that the intelligent machine will be human dependent. In other words, the human will remain in the centre of CPPS/CPS, and in I4.0.

The methodology to support this claim is inductive (based on informal induction methodology), similarly to the methodology applied in a number of texts found in literature. Therefore, considering limitations of the inductive methodology, the paper is a kind of a position paper, arguing that the human will be inevitably in the centre of the technological development, including AI/ML based development, referring as well to CPPS/CPS and I4.0.

The paper is divided into two parts. In the first part a review of argumentation from literature in favor of and against the thesis on the human role in future is presented. In addition, a concept of the *manufacturing singularity* is introduced, as well as an *intelligent machine architecture* towards the manufacturing singularity is presented, arguing that the intelligent machine will be always human dependent, and that, concerning the manufacturing, the human will remain in the centre. The argumentation is based on the phenomenon related to computational machine learning paradigm, as intrinsic feature of the AI/ML, through the inductive inference based machine learning algorithms. In the second part, an architecture of the Cyber-Physical (Production) Systems with multiple learning loops is presented, together with a set of experiments demonstrating the indispensable human role. Furthermore, the problem from the manufacturing community point of view on future of human role in Industry 4.0 as the environment for advanced AI/ML applications is discussed.

Considering the structure of this first part – Part I – in the Chapter 2, the actual scientific, and popular, discourses concerning coming massive use of AI/ML, referring to expectations of total automation, are briefly presented. The Chapter 3. introduces the concept of *manufacturing singularity*. Further, the Chapter 4, presents an *intelligent machine architecture*, as a model towards manufacturing singularity, and arguing that AI/ML will not overcome human based on computational requirements of inductive inference based machine learning. The argumentation is supported by the nature of a machine learning algorithms based on inductive inference, which effectiveness is provided only by the human. Finally the Conclusions of this first part – Part I – are given.

[4] For manufacturing industry, virtually more appropriate term would be Cyber-Physical Production Systems (CPPS). However, for short, we will use the term CPS as more general, without losing any CPS particular feature concerning manufacturing.

2. SCIENTIFIC AND POPULAR DISCOURSES CONCERNING COMING MASSIVE USE OF AI/ML AND EMERGENCE OF SUPER-INTELIGENCE

In many popular, as well as scientific discourses, it is suggested that the "massive" use of AI/ML, and especially its 'substantial progress' in so-called Artificial General Intelligence (AGI), and towards Artificial Super-Intelligence (ASI), will generate substantial changes, up to total exclusion of humans from work and decision making, even generating the 'existential risks' for humanity (a basic definition and a resume, as an initial information for the "newcomers" to the issue, could be consulted in e.g. [1] [5]).

The terms Artificial General Intelligence (AGI), and Artificial Super-Intelligence (ASI) nowadays are already widely accepted by the scientific community in the area as well as by the public in general through a numerous articles in public media (newspapers, magazines, popular journals, etc.).

Actually, the present state of the AI development is called Artificial Narrow Intelligence (ANI), sometimes called as well as *weak AI*, denominating the AI that could surpass the human, i.e. it could perform as a superintelligence, but only in a narrow domain, e.g. in chess playing[6], or, e.g. in many specific applications in manufacturing, but totally impotent in other areas.

From the other side, the Artificial General Intelligence (AGI), sometimes called as well as *strong AI*, or *Human-Level AI*, denominates the AI that is as smart as a human in any domain, hence "general", i.e. the AI that "can perform any intellectual task that a human being can" [2].

Finally, the Artificial Super-Intelligence (ASI), as defined by N. Bostrom [3], is "an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills". Artificial Super-Intelligence "ranges from a computer that's just a little smarter than a human to one that's trillions of times smarter" [2] than human as individual or collective.

---

[5] Although Wikipedia should not be considered as a source with scientific credibility (although the attribute '*scientific*' itself is also under the investigation), there are many entries that provide a very good introduction to basic concepts and terminology. One such entry is the entry on "*Existential risk from artificial general intelligence*" [1], (https://en.wikipedia.org/wiki/Existential_risk_from_artificial_general_intelligence), where the hypothesis on AI progress is formulated as:

*"**Existential risk from artificial general intelligence** is the hypothesis that substantial progress in artificial general intelligence (AGI) could someday result in human extinction or some other unrecoverable global catastrophe.[1][2][3] … If AI surpasses humanity in general intelligence and becomes "superintelligent", then this new superintelligence could become powerful and difficult to control. Just as the fate of the mountain gorilla depends on human goodwill, so might the fate of humanity depend on the actions of a future machine superintelligence [4.]"*

*[1] Russell, Stuart; Norvig, Peter (2009). "26.3: The Ethics and Risks of Developing Artificial Intelligence". *Artificial Intelligence: A Modern Approach*. Prentice Hall. ISBN 978-0-13-604259-4.

*[2] Bostrom, Nick (2002). "Existential risks". *Journal of Evolution and Technology*. 9 (1): 1–31.

*[3] Turchin, Alexey; Denkenberger, David (3 May 2018). "Classification of global catastrophic risks connected with artificial intelligence". *AI & Society*. 35 (1): 147–163. doi:10.1007/s00146-018-0845-5. ISSN 0951-5666.

*[4] Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies* (First ed.). ISBN 978-0199678112.

* the references in the Wikipedia's referred entry

[6] In May 1997, Deep Blue supercomputer defeated then world chess champion G. Kasparov 3½–2½

The question is whether the AI/ML will overtake humans in decision making and act without human intervention, including the decision making in general (in any domain), and particularly for the domain in manufacturing, which is of the manufacturing community interest.

It is suggested that it could happen when the AI/ML will achieve AGI and ASI levels. However, the opinions are divided. Within the technical and scientific engineering community in general, the computer science engineers and scientists, are looking much more far than others, i.e. looking much more far beyond I4.0, which, the I4.0, is usually the manufacturing engineering community horizon. And a part of that community, the computer science engineers and scientists, has a very dark vision.

In literature, when addressing the issue of threat by the (super) AI/ML, the well-known statements in recent years, by two famous public personalities, related to science and technology, Stephen Hawking (physicist) and Elon Musk (entrepreneur in the area of technology), that warn us about 'existential' threat, are referred almost regularly.

In an interview to BBC in 2014 Stephen Hawking said "The development of full artificial intelligence could spell the end of the human race" [4], while Elon Musk, in his interview, addressing to the National Governors Association 2017 Summer Meeting, stated that "AI is a fundamental risk to the existence of the human civilization …" [5].

Similar view, if not equal, are shared, or at least consider as a feasible hypothesis, as well by a number of scientists or professionals in the area of AI/ML. Some of the most popular are statements by, e.g.:

Russell S.J. and Norvig P. (AI scientists) in their very popular textbook on AI [6], referred to the hypothesis of that "The success of AI might mean the end of the human race" (Chapter 26.3, p. 1036), that could happen whether by AI/ML gained total autonomy or by human intention/error[7]. Bostrom N. wrote also a very popular book introducing Super-intelligence [3], in which he discussed in many details scenarios of the superintelligence development "paths, dangers (and) strategies" to deal with the superintelligence.

It is worth to refer to other books and papers that advocate the possibility to achieve AGI/ASI, such as [7–11], as well as evaluate the associated risks [12–30] and different strategies, policies and approaches how to deal with it [31–40].

It could be noticed that some of the authors referred above have a very high confidence, or even certainty, a fact, that AGI/ASI will come soon or later, and the associated issue of 'existential risk′, while other, virtually a majority, following a common positivistic scientific discourse practices, the threat by the superintelligence to the "existential risk", address more as a hypothesis, than as a (proven) fact (which is not).

The issue of "existential risk" is a consequence of the expected phenomenon related to AGI and ASI, the phenomenon so-called *singularity*, or *technological singularity* (as it is to be achieved by technological means). The notion of *singularity*, by literature sources, was introduced by V. Vinge in 1993 [41]. He wrote:

"We humans have the ability to internalize the world and conduct 'what-ifs' in our heads; we can solve many problems thousands of times faster than natural selection could.

[7] The main question considered in this paper is if there will be total AI/ML autonomy or not, which refers to the first part of the hypothesis, while the part of the hypothesis "the end of human race" by human intention/error already implies not possibility of the AI/ML total autonomy, or in other words, the human control of AI/ML.

Now, by creating the means to execute those simulations at much higher speeds, we are entering a regime as radically different from our human past as we humans are from the lower animals. From the human point of view this change will be a throwing away of all the previous rules, perhaps in the blink of an eye, an exponential runaway beyond any hope of control. Developments that before were thought might only happen in 'a million years' (if ever) will likely happen in the next century. We think it's fair to call this event a singularity ('the Singularity' for the purposes of this piece). It is a point where our old models must be discarded and a new reality rules. As we move closer and closer to this point, it will loom vaster and vaster over human affairs until the notion becomes a commonplace. Yet when it finally happens, it may still be a great surprise and a greater unknown."[8]

In other words, the *singularity* is a 'zone' beyond the point when development reach the full AGI, and continues autonomous development at the exponential rate, achieving ASI that would overpass human intelligence by 'trillions' of times, e.g. see [2], [9]. Similarly as in mathematics, where the singularity is a point in which e.g. a given function is not defined, or not possible to analyse, differentiates, the 'zone' of the singularity, in the context of AI/ML, the *singularity* is characterized by uncontrollable, autonomous development of AGI towards ASI, without intervention by human, and without possibility of intervention by human. Thus, this uncontrollable development could go in direction against humanity. *Singularity* is also described as "intelligence explosion", i.e. as "recursive self-improvement rapidly leading to superintelligence" [10] (similarly to the "explosion" of a mathematical function, e.g. $y=1/x$, near 0). Singularity is regularly referred issue in the publications on AGI/ASI, including the references abovementioned.

With the singularity another concept is closely related, the so-called *Fermi paradox*, named after famous Italian physicist Enrico Fermi (1901–1954). The Fermi paradox originally "refers to modern science's surprising failure to detect extraterrestrial life, provides evidence regarding the likelihood of the human species surviving long enough to become spacefaring" [29]. In the context of AGI/ASI, the Fermi paradox is considered in several ways referring to the problems of 1) the existence possibility, 2) friendliness or unfriendliness towards humankind, and 3) controllability, of AGI/ASI (see more in [30]). An important issue is the recognition of the moment when the AI/ML will turn to the *singularity*, i.e. the failure to recognize that moment (similarly as the "failure to detect extraterrestrial life"), and eventually to undertake some protective measures.

For the sake of "completeness", it is worth to mention that the development of AI/ML is not the only path to overcome the humans. Another path is through medical, or bio-technological, interventions substituting and upgrading our organs, by biological or nob-biological means, that includes e.g. uploading software in the brains, creating *cyborgs*, but also transcending it doing "even better by eliminating the human body entirely and uploading minds, creating a whole-brain emulation in software. Such an upload can live in a virtual reality or be embodied in a robot capable of walking, flying, swimming, space-faring or anything else allowed by the laws of physics, unencumbered by such everyday concerns as death or limited cognitive resources" [10]. This direction are called

[8] Concerning the first use of the term *singularity*, Vinge V. himself had referred, in the same paper [96], that "Von Neumann even uses the term singularity, though it appears he is still thinking of normal progress, not the creation of superhuman intellect."

transhumanism, and *posthumanism*, where *transhumanism* is an intermediate phase towards the *posthumanism*. The projects on brain modelling, and on connections of hard-ware/software to brain through implants and/or other means for 'brain-computer interfaces' or 'bio-digital fusion', are widely known in engineering community. For more details see e.g. [9, 10, 42], as further discussion on this topic is out of this paper's scope.

Nevertheless, both directions consider, at the end, exclusion of human, and, therefore, making the issue of threat to human existence, i.e. the "existential risk", and in the case of our paper, the problem of will the human be totally excluded from decision making in manufacturing.

The issue, of the "existential risk" and, in general, of advancing in AI/ML, has gained so wide impact that the issue was, and is, the subject of interest not only for the academic, research, innovation, business and engineering community, but it was also perceived as an important issue and became the subject of analysis and different political and policy measures, including concerns of national security, by the political institutions. Notable examples are President Obama's Executive Office's "two reports that laid out its plans for the future of artificial intelligence" [40], [43], Congressional Research Service's report [44] and, e.g., the European Parliament's Directorate General for Parliamentary Research Services issued a document titled "Should we fear AI" [45].

The 'existential risk' issue, by AI/ML, and how will be our future, especially in the context of the human's position, has gained also a great interest for the wider community, as well as for the public in general, and it became a theme in a number of leading magazines on science and in general public magazines, e.g.

The magazine "Nature" in April 2016 [46], dedicated the Editorial to the issue AGI/ASI, titled "Anticipating artificial intelligence", writing "Machines and robots that outperform humans across the board could self-improve beyond our control – and their interests might not align with ours. This extreme scenario, which cannot be discounted, is what captures most popular attention. But it is misleading to dismiss all concerns as worried about this."

In the same year, in September, the magazine "Scientific American" published an article by Christof Koch "Will Artificial Intelligence Surpass Our Own? – A philosopher worries about computers' ever accelerating abilities to outpace human skills" [47].

The magazine "TIME" has issued an extended Special edition entitled "Artificial Intelligence – The Future of Humankind" [48]. The literature and popular culture followed these concerns, if not led them, especially in the widely known movies[9].

However, these "views" are not new.

The views that machines will overcome humans, started already in 'late nineteenth century', mainly driven by Darwin's theory of evolution, combined with explosion of (mechanical) machines of all kinds, inherent to the (1st) industrial revolution. Actually,

[9] Just to remember the well-known plays/movies, such as:
"R. U. R." (Rossumovi Univerzální Roboti (Rossum's Universal Robots)), (1920) by Karel Čapek
"2001: A Space Odyssey" (1968) Stanley Kubrick
"Blade Runner" (1982) Ridley Scott
"Terminator" (1984) James Cameron and Gale Anne Hurd.
"The Matrix" (1999) the Wachowskis
"I, Robot" (2004) Alex Proyas

the visions on machines overcoming humans are based on analogical thinking, making analogy between humans and machines as organisms and their evolution. Similarly to biological evolution the machine evolution was envisioned as a process of making machines by machines, i.e. by "self-reproducing and evolving machines" (which is very close to nowadays research on self-reproducing robots and machines)

A good review of early ideas, from the 'late nineteenth century', is given in [49].

Virtually, the earliest (or one of just few earliest) author who wrote about the overcoming the humans by machines was Samuel Butler, in his essay, from 1863, "Darwin Among the Machines" [50]. In [50] Butler wrote: "*The upshot is simply a question of time, but that the time will come when the machines will hold the real supremacy over the world and its inhabitants is what no person of a truly philosophic mind can for a moment question."*

The 'machines will ultimately become' "the acme of all that the best and wisest men can ever dare to aim at" and that "man will have become to the machine what the horse and the dog are to man" (cited in [48]). The idea was further developed in much more details in famous "Erewhon" (1872) [51], later cited by A. Turing.

However, the ideas of Butler virtually do not have too much with AI/ML idea, and even less with the problem of AGI/ASI, i.e. with digital constructs (software), as his ideas were simply based on purely mechanical machines, considering the technological context of his time, which, the mechanical machines are simply much more limited in design than digital constructs, simply of their physicality. So, the Butler's ideas were purely "mechanistic", and mechanistic on a pure mechanics phenomenology, not considering even more abstract mechanistic phenomenology that could characterize software too (meaning conceiving software following so called Newtonian thinking, characterized by cause-effect phenomenon). Nevertheless, Butler is regularly referred as the earliest who suggested overcoming humans by machines, which is also the thesis which we are faced today.

In modern time, considering 'modern' in the context of beginning of computer science, and on the primordial beginning of AI/ML and further AGI/ASI, virtually the first who suggested the idea of overcoming humans by machines, or by AI, was Alan Turing. In 1951, Turing wrote, in the famous article titled "Intelligent Machinery, A Heretical Theory" [52] (reprinted in [53]): "*... at some stage therefore we should have to expect the machines to take control, in the way that is mentioned in Samuel Butler's 'Erewhon'*".

Later, in 1965, in one of the most referred paper concerning the AGI/ASI development, Irving John Good discussed the hypothesis of creation and consequences of an "ultraintelligent machine". He wrote:

"Let an ultraintelligent machine be defined as a. machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of man would be left far behind (see for example refs. *[cites three of his earlier papers]*). Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously." [54].

Of course, these concerns, and warnings, are not shared by all.

The list of critiques of possibility of AGI/ASI is long. Many of the texts already referred above list more or less extensively the critiques, i.e. the reasons why AGI/ASI, and the singularity associated, could not happen. For example, in [9] a relatively complete list is presented.

It is interesting a statement by A. Turing (which apparently contradicts to his statement abovementioned, but only if taken without the context): *"The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion."* [55] and (a few paragraph before the former citation) *"that the question, 'Can machines think?' should be replaced by 'Are there imaginable digital computers which would do well in the imitation game?'"* [55] (reprinted in [53]).

This Turing's statement is many times cited by Noam Chomsky in his numerous interviews and lectures, that could be found on YouTube[10], when asked on his opinion on possibility of AI, especially in the context on possibility of singularity. It is interesting that Chomsky, as one of the most famous intellectual of our time, is in total opposition to the other two famous personalities referred above, Hawking and Musk, concerning the possibility of AGI/ASI/singularity.

Another argument against, that should be accounted, came from another famous scientist, American philosopher, John Searle, who conceived the test called 'Chinese room' [56][11]. The 'Chinese room' is a "thought" argument (or test) that shows that computer program, regardless how intelligent or human-like it is, cannot have a "mind", "understanding" or "consciousness"'. The experiment could be shortly described as follows: a men, who do not knows Chinese, "either written or spoken", is closed in a room and have only a book of instructions on questions/answers on Chinese and a blank paper. The room has two "slots", one for receiving the questions on Chinese and the second to return the answers on Chinese too. The man in the room use the book with detailed instructions how to recognize the Chinese characters in questions and how to write the answers in Chinese. When we return the answers, the external observer, who gave the questions, can have impression that the man in the room knows and understand the Chinese, which is not the case. The man in the room can produce answers, using the book with the instructions, without any understandings of Chinese. Actually, it could be said that the 'Chinese room' represents a simulation model for the computer program for translations.

Searle introduced the distinction between two types, or levels, of AI, *weak and strong AI*[12]. He characterized the *weak AI* as AI that can only simulate understanding of Chinese, while the strong AI is that "that the programmed computer understands the stories and that the program in some sense explains human understanding" [56]. Here we will cite a paragraph from the abstract of his paper that summarize the conclusions:

""Could a machine think?" On the argument advanced here only a machine could think, and only very special kinds of machines, namely brains and machines with internal

[10] YouTube (2013) „Noam Chomsky on AI: The Singularity is Science Fiction!", https://www.youtube.com/watch?v=0kICLG4Zg8s&t=349s ; YouTube (2015) „Noam Chomsky: How dangerous is Artificial Intelligence?", https://www.youtube.com/watch?v=QOCQAtwdKqQ ; YouTube (2017) „Noam Chomsky - Can Machines Think?", https://www.youtube.com/watch?v=Ex9GbzX6tMo

[11] There could be easily find on the internet inumerous descriptions and figures of the „Chinese room" experiment.

[12] The terms are already introduced in this paper at the beginning of the Section 2.

causal powers equivalent to those of brains. And that is why strong AI has little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking." [56].

3. "MANUFACTURING SINGULARITY"

Speaking in terms of manufacturing systems, the question is whether the AI/ML will achieve *total* automation with exclusion of human from any single decision, or not. In fact, the question is the same as in the Chapter 2. Limiting the question to the area of manufacturing does not reduce the generality considered in the Chapter 2. It is obvious that the AGI/ASI implies exclusion of human from manufacturing as well. It means that, if we would like to discuss only the domain of manufacturing, i.e. the exclusion of human form decision making in manufacturing systems only, we will discuss only the *narrow AI* (ANI), or *weak AI* reach for the manufacturing systems domain.

At the one extreme, the *upper* extreme, the scenario is achieving *total* automation, without human, and at the other extreme the scenario is that there will be not any significant effects on the actual work, organizational and economic structure. The second extreme, the *lower* extreme, represents a trivial scenario, i.e. the state 'as is', with small and local optimizations. This extreme is not interesting for further examination. It means that we are interested in any scenario apart from the *lower* extreme, including the state of the *upper* extreme.

Concerning the first extreme, the question could be formulated, similarly as in the case of AGI/ASI, as if there is *singularity*. However, this could not be true singularity as defined in the Chapter 2, as here the 'singularity´ refers only to manufacturing, in the context of the extreme we are talking about, we will conditionally called it ***manufacturing singularity***.

This '*manufacturing singularity*' could mean the point of development of AI/ML for manufacturing, after which the program will become totally independent of human, including self-evolving. It means, that after that point of singularity, the program could autonomously make decisions on product design, including product variations, production, including decisions on production series, scheduling, distribution, but also design of new machine tools, manufacturing technologies, etc., and autonomously analyse the market, in terms of human population requirements, individual requests – implying understanding natural language (and imprecise) specifications. However, the manufacturing singularity would include surely the capability of the program improvement by itself, in a recursive processes with positive feedback, i.e. through multiple learning processes, and, of course, using 'big data', in real or 'batch' time. In other words, the *manufacturing singularity* means the program would be self-sufficient in all functional dimensions, including self-evolving.

The *manufacturing singularity* could be considered as a kind of *narrow*, or *weak*, singularity. The immediate question is if it is possible to call it 'singularity' if it is limited to a narrow domain? The positive answer to the question is based on the considerations abovementioned, Chapter 2, that "Artificial Narrow Intelligence (ANI), called as well as

weak AI, could surpass the human, i.e. it could perform as a superintelligence, but only in a narrow domain", hence *narrow*, or *weak*, singularity.

If the previous considerations are related to the definition, the following question is: "is *manufacturing singularity* possible at all", especially without general singularity, i.e. is it possible, phenomenologically, that the *manufacturing singularity* will keep itself within the manufacturing domain boundaries.

Considering the manufacturing system context, and its form of Cyber-Physical System (CPS) as an Industry 4.0 construct, the abovementioned question could take e.g. the following form:

> *Could it happen that developments in AI and ML make that AI and ML based* `{manufacturing system | CPS}` *undertake the production control from human, and in particular from the company owner, and start to produce on its own?*

This question is open, of course, similarly as the question on general singularity, as we saw in the Chapter 2. However, this question is not too relevant at this moment - remaining as the question for future research, the concept of *manufacturing singularity* could represent a reference concept, i.e. how close we can reach, or how much we can approximate to it.

It is not necessary to say that, if the *manufacturing singularity* is possible, whether as alone system or within the general singularity, and if the things would go in wrong direction, it would mean total destruction of the economy and social structure as we know it today. Of course we don't want it. So, we are strongly interesting in the role of human, does the human remain in control, or the human will be totally excluded. Thus, if we talk about approaching to, i.e. about approximation of, the *manufacturing singularity*, the problem we have is how to measure it in relation to the reference 'model'.

Fortunately, we have already an indirect, and simple, measure, or indicator. If the *manufacturing singularity* implies total substitution of human, this substitution has its manifestation in total elimination of actual organizational framework, heavily dependent of human working places, work forces, or jobs, in actual manufacturing sector. Therefore, the most simple and as the first measure we will use to evaluate approaching to, i.e. approximation of, the *manufacturing singularity*, will be to evaluate how much the AI/ML will affect, and in the limit elimination of, human working places, or jobs.

4. AN INTELLIGENT MACHINE ARCHITECTURE – TOWARDS *MANUFACTURING SINGULARITY* – AI/ML WILL NOT OVERCOME HUMAN

4.1. PRELIMINARIES

In this Chapter, an argument, in favour that it is not likely that human will be excluded from the decision making, is presented. The argument is based on the phenomenon related to computational machine learning paradigm, as intrinsic feature of the AI/ML.

This argument is developed through 1) presentation of the features of the machine learning algorithms based on inductive inference, 2) presentation of a concept for

an intelligent machine architecture, based on the phenomenology of the inductive inference based machine learning algorithms, and 3) experimentation with the synthesizing, or learning, of the manufacturing cell control program.

We have to say that there is a counter-opinion by V. Vinge [57] (reprinted paper [41] with annotations). He wrote that this (the argument based on computational complexity) is "the strongest argument against the possibility of the Technological Singularity", but to the reductionists it may "appear as a failure to solve the problem of software complexity. Larger and larger software projects would be attempted, but software engineering would not be up to the challenge, and we would never master the biological models". These conclusions are based on actual hardware technology which is limited in the context of the requirements. The progress is to be based on "large software projects and our progress in applying biological paradigms to massively networked and massively parallel systems", which would eliminate the argument we base on.

However, by our knowledge, this expectation is not yet supported by the algorithm theory.

4.2. MACHINE LEARNING ALGORITHMS BASED ON THE INDUCTIVE INFERENCE PARADIGM, AND THE ROLE OF HUMAN

There are many paradigms of learning, which could be classified into two large groups. The first group belongs to cognitive sciences and this group is out of our interest in this paper. The second group originated from computer sciences and consists of the so-called *machine learning* (ML) paradigms and algorithms. Although this classification is not exclusive, meaning that in both groups we found use of approaches from other group, all approaches that could be interpreted belonging to other group, will be understood in the context of machine learning.

Inductive inference, more precisely computational inductive inference, is a part of the area of machine learning, and it is one of the approaches to machine learning. In its basic form it belongs to the family of so-called *supervised learning*. However, it could be combined with other two big families: *unsupervised learning* and *reinforcement learning*. In a broader interpretation both *unsupervised learning* and *reinforcement learning* could be interpreted as a kind, or specific models of, inductive inference, or "Model Induction - Techniques to infer an explainable model from any model as a black box", see e.g. [58] (Fig. 2 in [58]). This includes also the popular, and powerful, *Deep Learning* models, which are the subset of the *Neural Networks* models, and which, further, could be modelled as a particular *representation class* in terms of inductive inference learning paradigm, e.g. see [59] (see below about the *representation class* concept, from [59]).

This is justified as all models take as input some training set. The role of human is always included in some way. (It is interesting to refer that many algorithms are considered as machine learning algorithms but by some other criteria they are not, e.g. pattern recognition algorithms are not).

Inductive inference could be interpreted as a true machine learning (different from e.g. "pattern recognition" – pattern recognition should not be confused with the pattern

learning). As an argument in favour of this claim could be taken a description by Valiant L. in [60], Chapter Seven: "*This chapter adopts Aristotle's dictum that beliefs come from two fundamental sources: syllogism and induction, or reasoning and learning. ... In previous chapters, I have distinguished subject matter that is theoryful (in the sense that an explanatory theory of it is known) from that which is theoryless, and I have argued that beliefs about the theoryless have the semantics of PAC learning because they are acquired, in the first instance, inductively by learning.*".

Similarly, in [61] the author gives the "methodological definition" of inductive inference, explicitly defining it as a machine learning form: "Inductive inference is automatic learning of formal grammars from finite subset of the language they generate."[13]. The learning process itself basically represents hypothesizing and checking the hypothesis against examples. By [62] "the term "inductive inference" denotes the process of hypothesizing a general rule from examples". In [59] (adopted from [63]) "programs that learn … (belong to) two broad categories … The first and more popular category falls under the label of artificial intelligence. Much of the work in this category tends to be less formal in direction and experimental in nature. The second category is termed inductive inference." Concerning the meaning of 'machine learning', we would cite the definition by Valiant L. in [63]: "In this paper we shall say that a program for performing a task has been acquired by learning if it has been acquired by any means other than explicit programming."

The seminal work by Valliant L. [63] provided the methodology to quantify the learnability of the algorithms, called *Probably Approximate Correct* (PAC) learning algorithms paradigm, conceived for learning concepts from a set of concepts, and with unknown distribution. The PAC learning paradigm "requires the learning algorithm to learn over all probability distributions *P*, even though the distribution is unknown. This is in keeping with our earlier observation that a good learner should learn with respect to any distribution, as long as the teaching and testing distributions are the same" [59]. These conditions are important as they give to the paradigm a generality and independence of any representation class used (see below for the "representation class"). The paradigm is called PAC because the learning algorithm outputs the "approximately correct" learned concept characterized by the learning error parameter ε, with the confidence parameter δ (see below for more details).

To evaluate the role of human within the learning process, i.e. within an intelligent machine, the outline of the learning algorithm is presented, following the paradigm of inductive inference based machine learning.

The symbols used, as well as the style of the presentation, are adopted from [59] and [63]. The inductive inference based learning model is presented as follows:

Let *f* be a *concept*, that is a subset of objects in predefined domain. A *class* of concepts *F* is any set of concepts.

Let *f* be the *target concept*, or the correct output concept, of the learning process.

The *inductive inference*, as a learning process about some concept, means that a learning algorithm takes, in an iterative process, as input a set of particular *examples* (one

[13] In [61] the author actually used the term "grammatical inference" instead of "inductive inference" but the methodology is strictly inductive, considering learning "from finite subset", meaning from examples, which is one of the main characteristics of the inductive inference.

by one) of the target concept f, by taking into account all given examples and by adding some new rules to build some hypothesis about the correct target concept (general rule) f.

Learning is provided by the separated learning algorithm which outputs the hypothesis about the target concept, or the learned concept, denoted by g, characterized by the parameters $\varepsilon \in [0,1]$ and $\delta \in [0,1]$, which are the learning error parameter ε, i.e. the error "allowed in a good approximation", and the level of confidence on δ, which "controls the likelihood of constructing a good approximation" [59]. The parameters ε and δ are typical features of the learning paradigm why it is called PAC. Both, f and g, belong to a corresponded class of concepts F.

Considering that our examples and the concept, the general rule, we want to learn are represented, in any case, in some language, we can develop the following definition:

> Let A be the alphabet and A^* is the set of all strings of finite length over A. A concept f is any subset of A^*. The set of examples given, denoted by I, is the set of sentences, which belong to the language L (concept) we want to learn. We can say that the inductive inference is attempt to find such grammar G that $I \subset L(G)$.

Each letter of the alphabet could be interpreted as a name of an (primitive) object, as, e.g., machine, program, or any other component of the manufacturing system about which we want to learn.

Basically, there are two types of examples, so called *positive* and *negative*. These will be represented by the pair (x, y), where x will be the string, sentence, representing a language (concept), and the $y=f(x)$ will be the indicator function, where $y \in \{1,0\}$. If $y = f(x) = 1$, then (x, y) is a *positive* example, representing the required language (concept) we want to learn, and if $y = f(x) = 0$, then (x, y) is a *negative* example, representing the language (concept) to be avoided.

Based on the previous considerations, the learning algorithm can be outlined, in the most simple form, as follows.

Learning algorithm outline (I):
input : the set (subset) S of F for learning
begin
pick a learning solution hypothesis $g \in F$ consistent with S;
output g;
end

Further detailing of the algorithm structure is as follows:

The size of S, i.e. the number of examples presented to the learning algorithm, should be of the polynomial dimension, meaning it cannot grow exponentially in our attempt to improve the learning outcome. Considering this condition, the learning algorithm will call a routine, e.g. called, EXAMPLE, which will provide m number of examples for learning (sample size). The number of examples m is in function of the following parameters: n, which is a maximum length of an example, the class of concepts F to be learned, and parameters ε and δ, which control the quantitative parameters of the learning process quality.

There are two additional requirements the algorithm should satisfy.

Considering the requirements for the algorithm tractability, i.e. the requirement for the polynomial-time *learnability*. It means that, in general case, the algorithms would be nondeterministic, for which the execution time will be expressed by a polynomial function. In other words, the algorithms should be efficient. Realisation of nondeterministic designing algorithms can be performed by introducing an additional source of information.

This additional source of information can be carried on by the routine which is called ORACLE. In a real system, the oracle can be a human expert, database, deduction system. It means that for hard problems we should include some expertise, heuristics, which will help in searching for the target concept. ORACLE "tells the learner whether or not the data positively exemplify the concept" [63].

An important requirement is the so-called, "representation class" of F. The "representation class" is an assignment of names for each concept f in F. The names are simply strings in Σ^*, (where Σ is an (binary) alphabet, and Σ^* is set of all strings of finite lenght on Σ), which is only in the interest of simplicity. We could choose the alphabet whic is more convenient [59]. The representation class is denominated R, and it could be said that we learn "F (in R)". The "representation classes" are in fact the "knowledge representation classes" we usually use in our research and engineering, e.g. conjunctive normal forms (CNF), disjunctive normal form (DNF), finite automata and regular languages, context-free languages, neural networks, networks in general, social networks, graphs in general, rational functions, splines, etc. Actually learning in a specific representation class implies specific algorithms as well, being all special cases of the PAC learning paradigm which is conceived over sets of concepts (independently of the concrete representation class) and with unknown distribution.

Considering the above referred requirements, the learning algorithm outline could be refined as follows:

Learning algorithm outline (II):
input : ε, δ, n, F *(in R)*
begin
let $m = f(\varepsilon, \delta, n, F)$;
make m calls of EXAMPLE;
let S be the set of examples seen;
    while a concept $g \in F$ is consistent with S **do**
    construct a concept $g \in F$ by calling **ORACLE**;
    end
output r /* $r \in R(g)$ *for some* $g \in F$;
end

The computational complexity of the algorithm, including the time complexity and the space complexity as well, is related (metaphorically) to the "Pick a learning solution hypothesis g" problem. In principle, it could be solved in two ways:

1) to generate the concept g randomly and to check its consistency, and
2) to construct the concept g accordingly with some strategy.

Construction of the concept *g* is a usual way of searching for a hypothesis of the target concept consistent with the sample (This process is related to the size of hypothesis concept space).

The appearance of the need for calling an ORACLE is critical. The argument in favour that it is not likely that human will be excluded from the decision making, by future development of AI/ML, is based on the need for an ORACLE. In fact, ORACLE implies human intervention.

In the case of engineering "intelligent systems", the role of the "oracle" is provided by domain expert, engineer. The theory shows that for some representation classes no calls of the oracle is needed, and from positive examples only, as e.g. for k-CNF, i.e. "conjunctive normal form expressions with a bounded number of literals in each clause" [63]. However, for more complex representation classes the calls for oracle are absolutely necessary[14].

It means that it is *not possible to perform the learning process of any more complex concept* ("hard" concepts in computational sense) *without human*. We could say that the oracle serves as a mechanism for the learning process "control strategy" definition, i.e. *the learning inference process management*.

In conclusion, *human is in the centre of learning*.

Some earlier *applications to manufacturing, of the approach presented, demonstrating always the role of human within the learning process*, could be consulted in [64–67]. In [68], an interpretation of the computational inductive inference base machine learning as a model of a Computational General Design Theory model is presented, demonstrating that the presented learning paradigm could be applied as the base for algorithmisation of the design as one of the most important functions of the *manufacturing singularity.*

4.3. AN INTELLIGENT. MACHINE ARCHITECTURE

Considering the relationship between the algorithms and machines, well defined in the theory of computation, formal languages and machines, see e.g. [69][15], we will use the term "machines" as an abstraction for an algorithm or its physical embodiment in a physical machine.

Learning a concept *f*, implies a separate learning algorithm, denoted here by *L*, that outputs the learned concept *g*. Our interest are manufacturing systems, or some of their components, e.g. manufacturing cell control program, *f*. The learning algorithm produces, or synthesizes, the learned concept of manufacturing systems, or some of its components, e.g. manufacturing cell control program, *g*. Both *f* and/or *g*, will be graphically presented by

[14] Also, for more complex classes, besides the absolute necessity for the calls of the oracle, both positive and negative examples are needed as well. The required properties of the oracle, in interpretation the structure of the human intervention, is also the subject of research, and there is a number of the oracle models. For more details see the numerous literature on the Computational Learning Theory, e.g. started with the already referred seminal paper by Valiant L. [63].

[15] The relationship between the languages and machines, i.e. between different classes of languages and corresponded classes of machines, could be found in virtually every textbook on the topic and on the internet, so, it will not be the object of further elaboration in this paper. This relationship is known as a Chomsky's hierarchy of languages and abstract machines, after the work by Noam Chomsky.

a block with input and output, following general system theory symbols, Figure 1.a. We call *f* and *g*, a system, or machine, *f* and system, or machine, *g*, as well. Similarly, the learning algorithm *L* will be also represented in the same manner, Figure 1b. Also, we can call *L*, a system, or machine, *L*. The relationship between the two systems *g* and *L*, will be represented graphically as in Figure 2. It is important to notice that output from *L* is not an input to *g*, but the output is the whole system *g*, i.e. the block part of the graphical representation of *g*. This is conventionally represented through the connection of the output arrow from *L* is on the bottom part of the *g* system´s graphics (keeping semantics that the input arrows are only connecting the left part).

We can say that the systems *g* and *L* belong to two levels, of which the system *g* represents the *object level*, we will denominate it as the *level 0*, while for the system *L* we will say to belong to the meta-level, we will denominate it as the *meta-level 1*. The reason to call the level of *L meta-level*, is that the system on the level 0 is the object of the system at the level 1, i.e. repeating, the system *L* does not present an input to the system *g*, but the system *L* outputs the system *g*. The architecture on the Fig. 2 represents, in fact, the architecture of an intelligent machine, corresponded to the *Learning algorithm outline (I)* (Section 4.2).

However, considering the subsequent requirements and their implementations, in the Section 4.2., especially the requirement for inclusion of the *oracle*, i.e. the requirement for inclusion of a *human* in the learning process, the architecture of the intelligent machine should be improved by inclusion of the *oracle*, i.e. *human*, denoted by *S*. To distinguish it from the 'mechanistic' part of the architecture, the oracle, i.e. human is represented by an ellipse, at the same level as *L*, as it is part of it. Therefore, as *L* is now composed of the oracle, i.e. human, and other 'mechanistic' part of the algorithm will be denoted as *L'*, Fig. 3.

Considering that *L*, *L'* and *S* belong to the *level 1*, i.e. to the *meta-level 1*. Their more complete denominations will be L_1, L_1' and S_1. The systems *L* and *g*, belong to two levels as shown in the architectures on the Figs. 2 and 3.

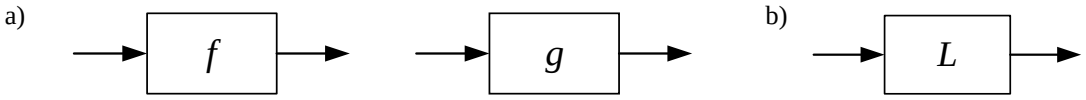

Fig. 1. Graphical representation of a) the learning *target concept f* for the learning process and the learning output (learned) concept *g*, and b) the learning algorithm *L*

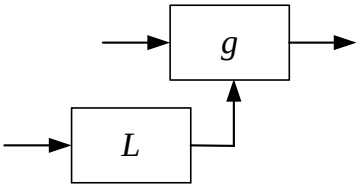

Fig. 2. Relationship between the learning algorithm (*L*) and the learned concept (*g*)

Further, considering that the learning algorithm L_1 could also be the object of learning by some other learning algorithm, meaning that L_1 is the output of another learning

algorithm L_2, we could introduce the second learning algorithm L_2, at the *meta-level 2*. The architecture could be further extended recursively, Figure 4, forming the architecture with multiple learning meta-levels.

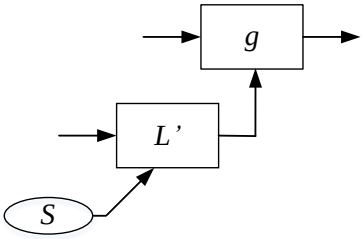

Fig. 3. Relationship between the learning algorithm (*L*) with *oracle* (*S*) and the learned concept (*g*)

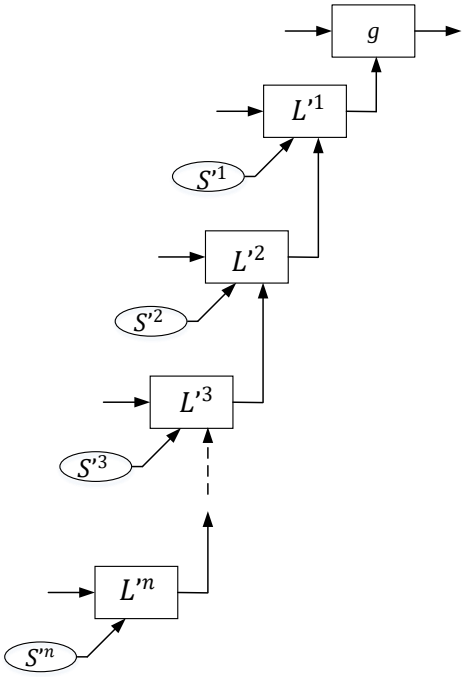

Fig. 4. Single intelligent machine architecture with multiple learning meta-levels

The machine, i.e. the system *f*, or *g* when learned, could be composed with other systems, or machines, as well as it may be a composition of its components. A sequential composition of various machines, or systems, whether *f*, or *g*, being components or compositions, is presented on Fig. 5.

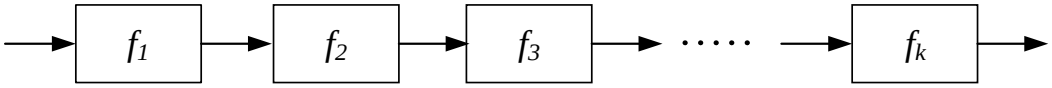

Fig. 5. Graphical representation of a sequential composition of various machines, or systems, whether *f*, or *g*,

Consequently, a total intelligent machine architecture is given on Fig. 6, as an adaptation of the intelligent machine architecture presented in [64].

As an algorithm is a deductive machine, the 'horizontal' composition of the object systems, or machines, *f* or *g*, at the *object-level 0*, represents the dimension of *deduction*.

On contrary, the 'vertical' composition of the learning algorithms, and moreover of the inductive inference algorithms, across the *meta-levels*, represents the dimension of *induction*.

As the inductive inference learning algorithms are also deductive machines, the induction is in fact *meta-deduction*, i.e. deduction about the deduction. Consequently, the induction dimension could be also denominated as dimension of *meta-deduction*.

Finally, the *oracle*, i.e. *human*, modules composition, being at the same levels as the corresponding learning algorithms, represent the 3rd dimension that will be denominated as the dimension of *learning control*.

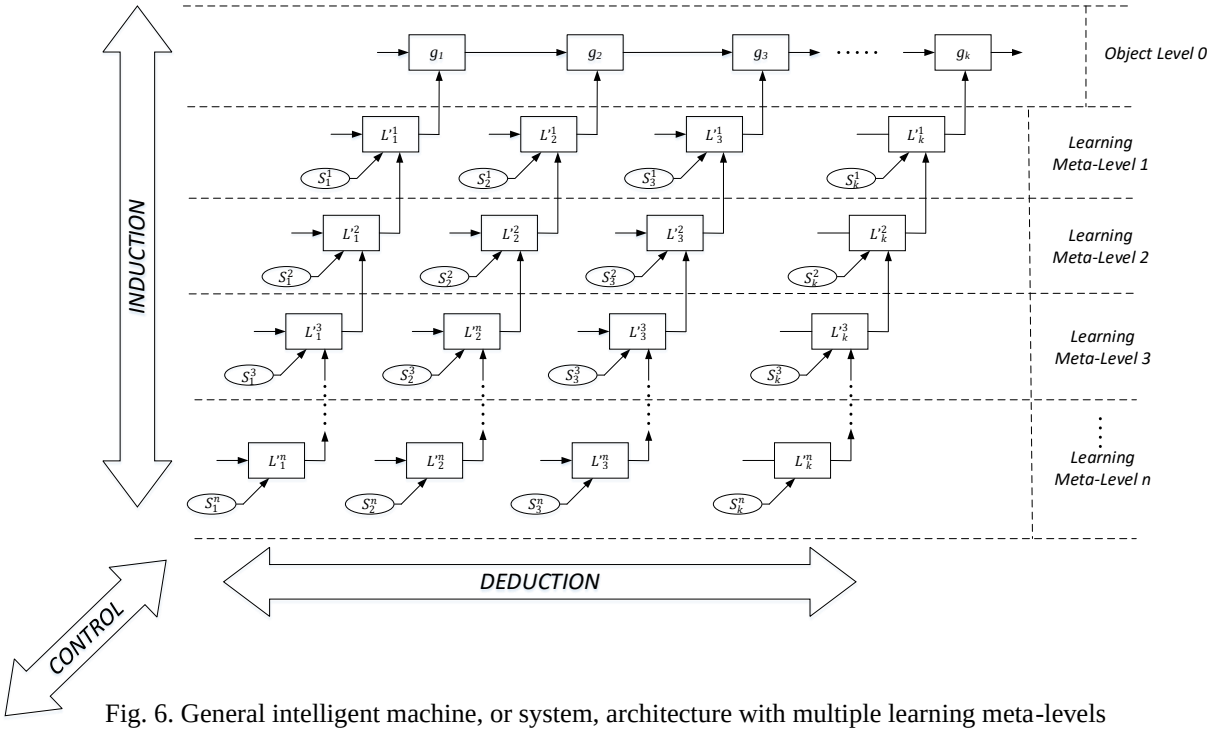

Fig. 6. General intelligent machine, or system, architecture with multiple learning meta-levels

Although the feedback relations, or loops, are the paradigmatic part of any modern engineering control system, for the high level of abstraction of the intelligent machine architecture conception and presentation, Figure 6, the feedback relations, or loops, are not relevant.

The feedback relations, or loops, are to be presented in the architecture models oriented to implementations, as, e.g. the Intelligent Cyber-Physical (Production) System logical architecture – see Part II of this paper, which follows philosophy of the intelligent machine architecture conception and presentation on Fig. 6.

To make a correspondence to some of the concept and terminology used in other scientific areas, especially to the theories of learning, but also useful to introduce in manufacturing area especially concerning intelligence, i.e. AI/ML embedding in manufacturing systems, we will introduce the terms of single- loop learning systems, double-loop,

and in general, n-loop learning systems. Considering the total intelligent machine architecture, Fig. 6, the special cases of the architecture containing only the first meta-level, i.e. containing only the first level of learning, Fig. 7a., will be denominated *single-loop* learning intelligent machines.

Similarly, the special cases of the architecture containing only two meta-levels, i.e. containing the first and the second level of learning, Fig. 7b., will be *denominated double - loop* learning intelligent machines. Whereas in the general case, the architecture containing *n* meta-levels, i.e. containing *n* levels of learning, Fig. 7c., will be denominated *n-loop* learning intelligent machines.

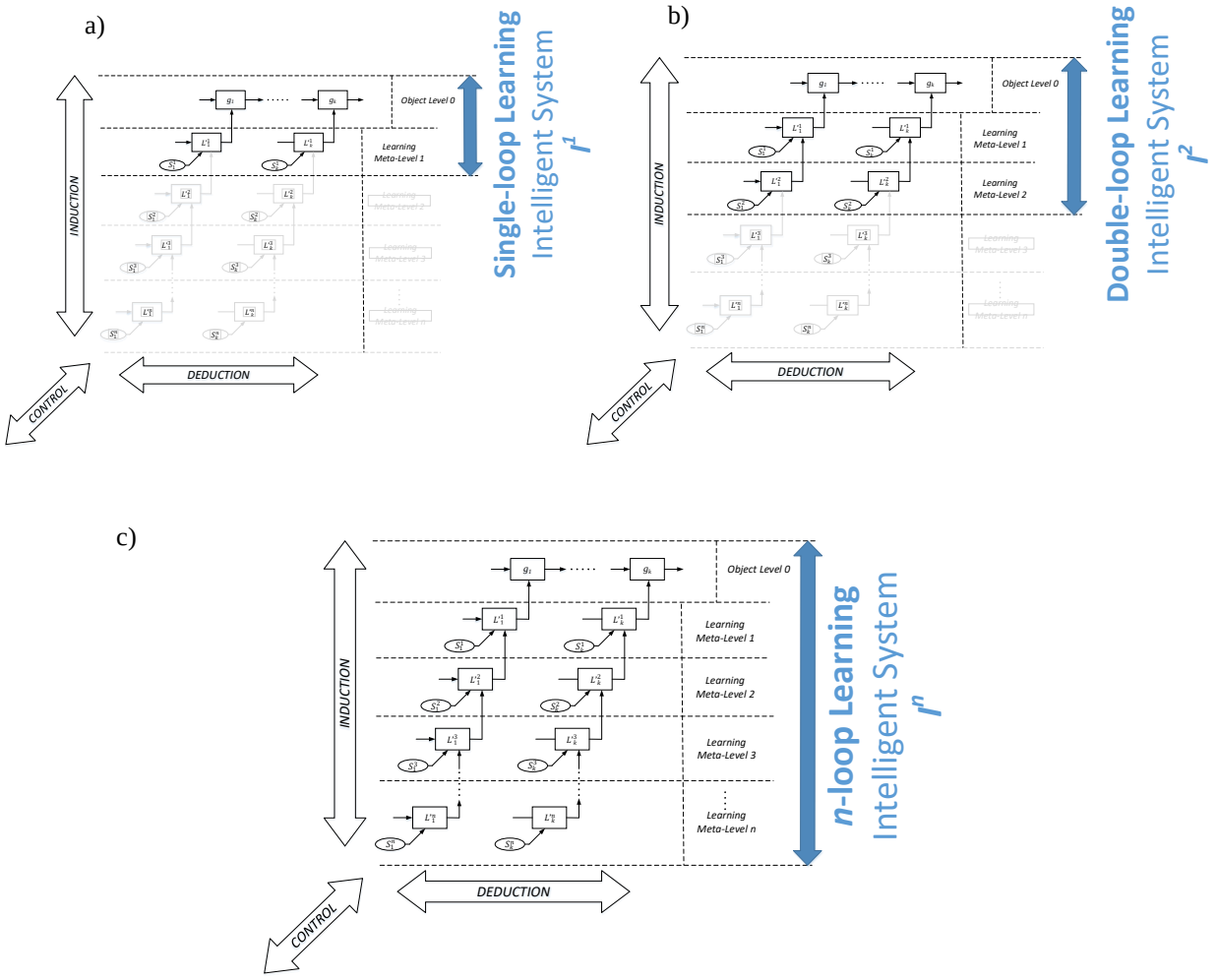

Fig. 7. General intelligent machine, or system, architecture with a) single-loop learning, b) double-loop learning, c) *n*-loop learning

5. CONCLUSIONS

The main theme of the paper is: does the new developments of AI/ML, and projected paradigms of AGI/ASI will totally exclude humans from decision making? In support to answer to this question a number of arguments from literature in favour and against are

reviewed. The authors' clear position against the possibility that the future development of AI/ML will exclude humans from the decision making is presented.

In this first part of the paper, three contributions are presented. First, the argument in support of the authors' position is presented. The argument presented by the authors is based on the phenomenon related to computational machine learning paradigm, as intrinsic feature of the AI/ML. This argument is developed through the presentation of the features of the machine learning algorithms based on inductive inference, which demonstrate that an effective learning algorithm depends on human intervention, or guidance, putting the human in the centre of the process.

There was referred a counter-argument, from the literature, suggesting that "applying biological paradigms to massively networked and massively parallel systems" would overcome the problems. However, by the authors' knowledge, this expectation is not yet supported by the algorithm theory.

Secondly, the concept of *manufacturing singularity* is defined, following the definition of the general singularity in the context of AGI/ASI development. There is a question if it is possible or not. However, independently of the answer, considering that this question is not too relevant at this moment and that remains the question for future research, the concept of *manufacturing singularity* could represent a reference concept, i.e. how close we can reach, or how much we can approximate to it.

Thirdly, a general intelligent machine architecture with multiple learning meta-levels is defined based on the phenomenology of the inductive inference based machine learning algorithms. This architecture is further used as a reference model for defining of an Intelligent CPS with multiple learning loops (I^n-CPS), presented in the Part II of this paper.

The future work, surely, should address further development of the features of the *manufacturing singularity*, as well as measures for evaluating approximation to the *manufacturing singularity*. Development of the intelligent architecture structure is also a task for the future, but these considerations are out of this paper's scope.

Other questions concerning the evaluation of the human performance during the learning process within the I^n-CPS, i.e. more precisely within the I^1-CPS, as well as evaluation of the AI/ML employment impact on manufacturing systems and industry is presented in the Part II.

ACKNOWLEDGEMENTS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020

REFERENCES

[1] Wikipedia, 2020, *Existential risk from artificial general intelligence*, https://en.wikipedia.org/wiki/Existential_risk_from_artificial_general_intelligence

[2] URBAN T., 2015, *The AI Revolution: the Road to Superintelligence*, Wait But Why, https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html.

[3] BOSTROM N., 2014, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press.
[4] CELLAN-JONES R., 2014, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC News, https://www.bbc.com/news/technology-30290540.
[5] MUSK E., 2017, *Elon Musk at National Governors Association*, 2017 Summer Meeting, https://www.c-span.org/video/?431119-6/elon-musk-addresses-nga&start=5049.
[6] RUSSELL S., NORVIG P., 2009, *Artificial Intelligence.* Artificial intelligence: A modern approach, 3rd edition, 2016, Prentice Hall.
[7] GOERTZEL B., PENNACHIN C., (Eds.), 2007, *Artificial General Intelligence*, New York, Springer.
[8] YAMPOLSKIY R.V., 2005, *Artificial Superintelligence: a Futuristic Approach*, CRC Press.
[9] KURZWEIL R., 2005, *The Singularity is Near: When Humans Transcend Biology*, Penguin.pdf.
[10] TEGMARK M., 2017, *Life 3.0: Being Human in the Age of Artificial Intelligence*, Alfred A. Knopf.
[11] EDEN A. H., MOOR J.H., SØRAKER J.H., STEINHART E., (Eds.), 2012, *Singularity Hypotheses: A Scientific and Philosophical Assessment*, Springer.
[12] MÜLLER V.C., BOSTROM N., 2016, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, Fundamental issues of artificial intelligence, Springer, 553–571.
[13] NILSSON N.J., 2005, *Human-Level Artificial Intelligence? Be serious*!, AI magazine, 26/4, 68–75.
[14] BOSTROM N., YUDKOWSKY E., 2014, *The Ethics of Artificial Intelligence*, Frankish, K., Ramsey, W. M. (Eds.), The Cambridge handbook of artificial intelligence, The Cambridge handbook of artificial intelligence, Cambridge University Press., 1, 316–334.
[15] LIU H.Y., 2018, *The Power Structure of Artificial Intelligence*, Law, Innovation and Technology, 10/2, 197–229.
[16] GOERTZEL B., 2013, *The Structure of Intelligence: A New Mathematical Model of Mind*, Springer Science & Business Media.
[17] GOERTZEL B., 2014, *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*, Journal of Artificial General Intelligence, 5/1, 1–48.
[18] TURCHIN A., DENKENBERGER D., 2020, *Classification of Global Catastrophic Risks Connected with Artificial Intelligence*, AI & SOCIETY, 35/1, 147–163.
[19] YUDKOWSKY E., 2008, *Artificial Intelligence As a Positive and Negative Factor in Global Risk*., Nick Bostrom and Milan M. Ćirković (Eds.) Global Catastrophic Risks, 308–345.
[20] BOSTROM N., 2002, *Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards*, Journal of Evolution and technology, Vol. 9/1.
[21] TORRES P., 2019, *Existential Risks: a Philosophical Analysis*, Inquiry, 1–26.
[22] BEARD S., ROWE T., FOX J., 2020, *An Analysis and Evaluation of Methods Currently Used to Quantify the Likelihood of Existential Hazards*, Futures, 115, 102469.
[23] BAUM S.D., 2020, *Quantifying the Probability of Existential Catastrophe: A reply to Beard et al*. Futures, 123, 102608.
[24] VINCENT C., MÜLLER., 2014, *Risks of General Artificial Intelligence* (Editorial), Journal of Experimental & Theoretical Artificial Intelligence, 26/3, 297–301.
[25] SOTALA K., YAMPOLSKIY R.V., 2015, *Responses to Catastrophic AGI Risk: a Survey*, Physica Scripta, 90/1, 018001.
[26] YAMPOLSKIY R.V., SPELLCHECKER M.S., 2016, *Artificial Intelligence Safety and Cybersecurity: A Timeline of AI Failures*, arXiv Preprint arXiv, 1610.07997.
[27] TORRES P., 2019, *The Possibility and Risks of Artificial General Intelligence*, Bulletin of the Atomic Scientists, 75/3, 105–108.
[28] ĆIRKOVIĆ M. M., 2015, *Linking Simulation Argument to the AI risk*, Futures, 72, 27–31.
[29] MILLER J.D., FELTON D., 2017, *The Fermi Paradox, Bayes' Rule, and Existential Risk Management*, Futures, 86, 44–57.
[30] PISTONO F., YAMPOLSKIY R.V., 2016, *Unethical Research: How to Create a Malevolent Artificial Intelligence*, https://arxiv.org/abs/1605.02817.
[31] CRITCH A., KRUEGER D., 2020, *AI Research Considerations for Human Existential Safety (ARCHES)*, arXiv Preprint arXiv:2006.04948.
[32] CALO R., 2017, *Artificial Intelligence Policy: a Primer and Roadmap*. UCDL Rev., 51, 399.
[33] WOGU I.A.P., 2017, *Artificial Intelligence, Alienation and Ontological Problems of Other Minds: A Critical Investigation Into the Future of Man and Machines*, 2017 International Conference on Computing Networking and Informatics (ICCNI) , IEEE, 1–10.
[34] BOYD M., WILSON N., 2020, *Catastrophic Risk from Rapid Developments in Artificial Intelligence*, Policy Quarterly, 16/1, 53–61.

[35] ČERKA P., GRIGIENĖ J., SIRBIKYTĖ G., 2015, *Liability for Damages Caused by Artificial Intelligence*, Computer Law & Security Review, 31/3, 376–389.
[36] RUSSELL S., DEWEY D., TEGMARK M., 2015, *Research Priorities for Robust and Beneficial Artificial Intelligence*, Ai Magazine, 36/4, 105–114.
[37] CASTEL J.G., CASTEL M.E., 2016, *The Road to Artificial Super-Intelligence: Has International Law a Role to Play*? Canadian Journal of Law and Technology, 14/1.
[38] Future of Life Institute, 2015, *An Open Letter – Research Priorities for Robust and Beneficial Artificial Intelligence*, Future of Life Institute, https://futureoflife.org/ai-open-letter/.
[39] European Parliament, 2018, *Should We Fear Artificial Intelligence*, European Parliament – Directorate-General for Parliamentary Research Services – Scientific Foresight Unit (STOA), ISBN 978-92-846-2676-2.
[40] AGRAWAL A., GANS J., GOLDFARB A., 2018, *The Obama Administration's Roadmap for AI Policy.* Harvard Business Review. https://hbr.org/2016/12/the-obama-administrations-roadmap-for-ai-policy
[41] VINGE V., 1993, *The Coming Technological Singularity: How to Survive in the Post-Human Er*a. Proceedings of VISION-21 Symposium. NASA Conference Publication 10129. Westlake, Ohio. http://www.frc. ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html.
[42] BOSTROM N., 2005, *A History of Transhumanist Thought*. Journal of evolution and technology, 14/1, 1–25.
[43] HOUSE, WHITE, 2018, *Update From the National Science and Technology Council Select Committee on Artificial Intelligence*, White House, Office of Science and Technology Policy.
[44] HOADLEY D.S., LUCAS N.J., 2018, *Artificial Intelligence and National Security*, Congressional Research Service.
[45] Should we fear AI – European Parliament-DG Parliamentary Res Services – Sc Foresight Unit (STOA) EPRS _IDA(2018)614547_EN.pdf.
[46] Nature, 2016, Anticipating Artificial Intelligence (Editorial), *Nature*, 532, 413.
[47] KOCH C., 2015, *Will Artificial Intelligence Surpass Our Own*, Scientific American. https://www.scientific american.com/article/will-artificial-intelligence-surpass-our-own.
[48] KENNEDY K., MIFSUD C., (Eds.), 2017, *Artificial Intelligence – The Future of Humankind*, TIME Special editions.
[49] TAYLOR T., DORIN A., 2018, *Past Visions of Artificial Futures: one Hundred and Fifty Years Under the Spectre of Evolving Machines*, Artificial Life Conference Proceedings 91–98, MIT Press.
[50] BUTLER S., 1863, *Darwin Among the Machines*, [to the Editor of the Press, Christchurch, New Zealand, 13 June, 1863.]. A First Year in Canterbury Settlement with Other Early Essays, 180-5. NZETC – New Zealand Electronic Texts Collection. http://nzetc.victoria.ac.nz/tm/scholarly/tei-ButFir-ButFir-f4.html.
[51] BUTLER S., 1872, *Erewhon*, Edition 1974, Penguin UK.
[52] TURING A.M., 1996, *Intelligent Machinery, a Heretical Theory (c. 1951)*. Philosophia Mathematica, 4/3, 256–260, https://doi.org/10.1093/philmat/4.3.256.
[53] COPELAND B.J., (Ed.), 2004, *The essential Turing*, Oxford University Press.
[54] GOOD I.J., 1966, *Speculations Concerning the First Ultraintelligent Machine*, Advances in computers, 6, 31–88, Elsevier.
[55] TURING A., 1950, *I.–Computing Machinery and Intelligence, Mind*, LIX(236), 433–460, https://doi.org/ 10.1093/mind/LIX.236.433
[56] SEARLE J.R., 1980, *Minds, Brains, and Programs*, The Behavioral and Brain Sciences, 3, 417–457.
[57] VINGE V., 2003, *Technological Singularity*, http://www8.cs.umu.se/kurser/5DV084/HT10/utdelat/vinge.pdf.
[58] GUNNING D., AHA D.W., 2019, *DARPA's Explainable Artificial Intelligence (XAI) Program*, AI Magazine. 40/2, 44-58, https://doi.org/10.1609/aimag.v40i2.2850.
[59] NATARAJAN B.K., 2014, *Machine Learning: a Theoretical Approach*, Elsevier.
[60] VALIANT L., 2013, *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*, Basic Books (AZ).
[61] MICLET L., 1990, *Grammatical Inference*, Bunke H., & Sanfeliu A., (Eds.) Syntactic and Structural Pattern Recognition – Theory and Applications, World Scientific, 237–290.
[62] ANGLUIN D., SMITH C.H., 1983, *Inductive Inference: Theory and Methods*, ACM computing surveys (CSUR), 15/3, 237–269.
[63] VALIANT L., 1984, *A Theory of the Learnable*, Communications of the ACM, 27/11, 1134–1142.
[64] PUTNIK G., 1993, *Application of the Inductive Learning Based on Automata Theory for Tooling Selection in Manufacturing Systems*, Dr.Sci. Thesis, Mechanical Engineering Faculty, University of Belgrade, Belgrade, Serbia, (in Serbian).
[65] PUTNIK G.D., ROSAS J.A., 1997, *LEARN – A Prototype Software Tool for Machine Learning*, Proceedings of the 2nd World Congress on Intelligent Manufacturing Processes and Systems, Budapest, (L. Monostori; Ed.), Springer, 587–592.

[66] PUTNIK G.D., ROSAS J.A., 1997, *Manufacturing System Simulation Model Synthesis: Towards Application of Inductive Inference*, L.M. Camarinha-Matos (Ed.) Reengineering for Sustainable Industrial Production, Proceedings of OE/IEEE/IFIP International Conference on Integrated and Sustainable Industrial Production – ISIP ′97, Chapman & Hall, 259–272.
[67] PUTNIK G.D., ROSAS J.A., 2001, *Manufacturing System Design: Towards Application of Inductive Inference*, Proceedings of the International Workshop on Emerging Synthesis – IWES 01, CIRP sponsored, Bled, Slovenia.
[68] PUTNIK G.D., 2011, *A Computational General Design Theory model as an interpretation of the Computational Inductive Inference*, (unpublished manuscript).
[69] DENNING P.J., DENNIS, J.B., QUALITZ J.E., 1978, *Machines, languages, and computation*, Prentice-Hall.