Towards Minimally Conscious Finite-State Controlled Cyber-Physical Systems

A Manifesto

Jiří Wiedermann

Jan van Leeuwen

Technical Report UU-PCS-2020-01
May 2020

Center for Philosophy of Computer Science
Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

Series: UU-PCS

Department of Information and Computing Sciences
Utrecht University
Princetonplein 5
3584 CC Utrecht
The Netherlands

Towards Minimally Conscious Finite-State Controlled Cyber-Physical Systems A Manifesto[*]

Jiří Wiedermann[1] and Jan van Leeuwen[2]

[1] Institute of Computer Science of Czech Academy of Sciences and Karel Čapek Center for Values in Science and Technology, Prague, Czech Republic
jiri.wiedermann@cs.cas.cz
[2] Dept of Information and Computing Sciences, Utrecht University, the Netherlands
J.vanLeeuwen1@uu.nl

Abstract. Incidents like the crash of Lion Air Flight 610 in 2018 challenge the design of reliable and secure cyber-physical systems that operate in the real-world and cope with unpredictable external phenomena and error-prone technology. We argue that their design needs to guarantee minimal machine consciousness, expressing that these systems must operate with full awareness of (the state of) their components and the environment. The concept emerged from our recent effort to develop a computational model for conscious behavior in robots, based on the theory of automata. Making systems minimal machine conscious leads to more trustworthy systems, as it strengthens their behavioral flexibility in varying environments and their resilience to operation or cooperation failures of their components or as a whole. The notion of minimal machine consciousness has the potential to become one of the defining attributes of Industry 4.0.

"We don't need artificial cognitive agents. We need intelligent tools."
D. C. Dennett [3], 2019

Keywords: automata theory, cyber-physical systems, design philosophy, Industry 4.0, minimal machine consciousness, self-control.

1 Introduction

Aircraft crashes like that of Lion Air Flight 610 and collisions of self-driving vehicles can often be reduced to combined failures in the hardware and software components of their underlying systems. Incidents like this seriously challenge the design of reliable and secure systems that operate in the real world and can cope with unpredictable external phenomena and error-prone technology. How should one look at the issues at stake here, from a design philosophical viewpoint?

Reliability and safety are crucial in *all* cyber-physical systems, not just in the examples we gave. Cyber-physical systems are systems of many components in which computers (processors) are used to govern the behavior of some parts of the physical world. The systems comprise both the computers and the parts of the physical world they govern [1, 6].

[*] Version dated: May 23, 2020. The research of the first author was partially supported by ICS AS CR fund RVO 67985807, programme Strategy AV21 "Hopes and Risks of the Digital Age", and the Karel Čapek Center for Values in Science and Technology.

Example 1. Examples of cyber-physical systems are ATM's, heart pacemakers, mobile phones, smart TVs, driverless cars, aircraft, trains, lifts, cranes, power plants, sea walls, ships, hadron colliders, orbital space stations, manufacturing systems, and many other systems. (Cf. [6].)

In fact, one may well see cyber-physical systems as generalized robots. Their development is rapidly progressing with the increasing use of advanced techniques from AI and yet, incidents like mentioned above continue to occur.

Generally speaking, one may argue that today's cyber-physical systems still operate as robot 'zombies' when it comes to adjusting to new or varying environments, their 'awareness' of operation and cooperation failures of their components or as a whole, and reacting properly to combined malfunctions of their modules. A natural question is whether the very use of insights from AI could ameliorate this state of affairs, possibly even drastically.

From a philosophical perspective, the vulnerability of cyber-physical systems is rooted in their lacking or limited cognitive abilities. Even if we do not know how to endow such systems with the facilities of full-blown intelligence or even consciousness, and perhaps we might not even want to build such systems (cf. [3]), we can imagine to equip them with important aspects of awareness and behavioral knowledge of the parts of the world that they perceive via their sensors.

In this paper we argue that the design of cyber-physical systems must guarantee, what we call, *minimal machine consciousness*, a concept expressing that the systems must operate and act based on, and maintaining, full awareness of (the state of) their components and the situation in their environment.

The concept of minimal machine consciousness has emerged in our recent effort to give a practical model for conscious behaviour in robots, based on the theory of automata [12]. The concept was initially meant to provide an exploratory, theoretical approach to the computational modeling of certain basic aspects of consciousness. However, we will show that the underlying ideas can *also* be used in industrial applications - namely in the design of reliable and secure cyber-physical devices operating in the real world.

We contend that designing cyber-physical systems to be minimal machine conscious is the key to obtaining trustworthy systems, as it strengthens their behavioral flexibility in new or varying environments and their resilience to operation or cooperation failures of their components or as a whole. As a design objective, the notion of minimal machine consciousness seems to provide the missing link to obtaining safe cyber-physical systems, and it should therefore be applied wherever possible and appropriate. This is summarized in the following *manifesto*:

> *All cyber-physical systems operating in a given environment, with or without human aid, must be designed as minimal machine conscious cognitive systems.*

In the remainder we discuss the essence of the design philosophy we propose. In Section 2 we outline the architectural basis of the cyber-physical systems that may be termed 'cognitive'. Then, in Section 3, we define the 'four principles' of minimal machine consciousness, and we argue why cognitive cyber-physical systems have all it takes to satisfy the criteria. The model derives from the general framework in [12] and relates to a refinement of the typical operational cycle of robotic systems.

In the subsequent sections we consider why minimal machine consciousness [12] gives us a proper tool for the design challenge we posed. In Section 4 we discuss what is typically made possible by minimal machine consciousness. We also discuss the potential for realizing cognitive cyber-physical systems and the potential of minimal machine consciousness for becoming one of the defining system attributes of *Industry 4.0*. Finally, in Section 5, we offer some conclusions.

2 Cognitive Cyber-Physical Systems

2.1 Architecture

In general, a cyber-physical system is an embedded entity of components that is producing a behavior in some environment, based solely on the inputs from its sensory and motor modules. Some systems also take inputs from human operators. See [1] for a general introduction to the foundations of cyber-physical systems.

In cognitive cyber-physical systems, specific conditions are imposed that allow a qualitative assessment of the information obtained from all sources of input and from the appropriate actions that result from it.

The architecture of a cognitive cyber-physical system C consists of four main parts: its sensory units, its motor units (or effectors), its finite-state control unit and its dashboard. See Figure 1 for a typical systems view.

Sensory units Sensory units are devices, modules, or subsystems whose purpose is to detect and register events or changes in the system's environment and send the information about them to the control unit of the system. A sensory unit (or sensor) sends both a *representation* of the occurrence of a phenomenon it is specialized to and, depending on the type of sensor, also a feedback signal representing the *accuracy* of the corresponding sensation. The accuracy of a sensation can be graded according to some scale (such as insufficient, low, fair, excellent, and so on) and depends on the nature of that sensation. For example, it can be its magnitude, intensity, frequency, blurriness, etcetera.

Motor units Motor units are devices, modules, or subsystems whose purpose is to perform one or more actions in the environment, seen as components of the system's behavior. Some motor units may serve for the positioning of sensors or of the system as an embodied entity, others may be designed for the manipulation of various effectors of the system. Motor modules send feedback to the control unit in the form of *reports* stating whether, or to what extent, the proposed operation could be realized. The feedback 'accuracy' and the 'reports' together are called the *quality* of the respective feedback. The qualities of the sensations and of the reports from the motor modules provide important feedback information for a system's 'self-monitoring' and 'self-awareness'. The graded responses allow the system to monitor the working of its sensory and motor modules. Clearly, not all effectors must perform mechanical movement. Some of them may be 'transmitters' that just produce internal or external signals of some kind: optical, chemical, acoustic, tactile, visual, radio-magnetic, etc. The emitted signals are used for communication purposes, under the assumption that the system and its environment possess receivers for the respective signals.

Finite-state control The finite-state control unit is the *computational* heart of any cognitive cyber-physical system. It acts in a similar fashion as deterministic finite-state automata. The control unit iterates the *operational cycle* of the system (see below). In any iteration, its purpose is to determine the (next) set of actions of the entire system based on four ingredients: the *current state* of the control unit, the current sensations from the sensory modules, the quality of these sensations, and the current reports from the motor modules. Typically, the finite-state control unit will be a *multiprocessor* that is programmed to generate the instructions for the set of actions to follow. States represent the possible *configurations* of the unit.

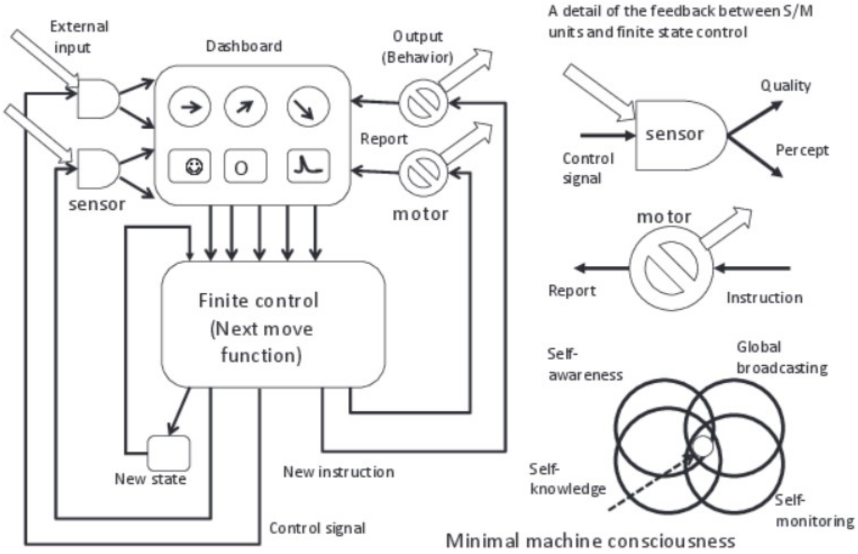

Fig. 1. A schema of a cognitive cyber-physical system

Presuming that there is but a finite number of sensory and motor modules and that the control unit can recognize but a finite number of signals of various types received from these modules, a control unit can produce but a finite number of different instructions. Each such instruction states for a specific sensory or motor module what it has to do ('in the next step'). The instructions may require repeating the previous action, or performing a new specific action, or doing nothing at all. The number of different actions can be very large, even exponential in the number of received signals.

Dashboard The four ingredients on which the control unit operates are jointly called the *dashboard* information of C. (At any time t, this information can be seen as the *instantaneous description* of the system at time t.) We note that all sufficiently complex (cognitive) cyber-physical systems have some form of 'physical' dashboard that has no influence on computation but is only used by a human operator for keeping a system's behavior within reasonable 'boundaries'. This human activity can be partially or fully automated, as we expect it is in self-controlled systems.

Note that in principle, the control unit works orders of magnitude faster than many other modules, especially the mechanical ones, of a cognitive system. Therefore the entire system works in an asynchronous manner.

2.2 Operation

Another characteristic feature of a cognitive cyber-physical system is its particular *operational cycle*. It is a variant of the well-known 'robotic paradigm', i.e. the *Sense-Think-Act* or *Sense-Plan-Act* cycle, now consisting of four phases that are iterated in sequel: *Sense-Analyze-Compute-Act* (SACA). The 'Analyze-Compute' part may be seen as a refinement of the 'Think' or 'Plan' phase in the standard robotic case. The four phases are distinguished as follows.

Sense In the first phase ('sense') dashboard information is retrieved, in parallel, by the control unit. The dashboard information must be read in parallel since the next proceeding of the system must be based on all available information at the time an iteration cycle begins.

Analyze In the second step ('analyze'), the dashboard information and its gradings are interpreted and fitted against the state information of the finite control, so as to determine how the system and its actions are progressing internally and, naturally, in the system's environment (as far as it can tell from its sensory input). The phase leads to a decision for a next action.

Compute In the third step ('compute'), a so-called *transition function* is applied to the dashboard information and the anticipated decision. This is a function that for any given current state, current dashboard information, and the current analysis of it, determines a new state of the control unit and for each sensory and motor module a new action to be realized in this iteration cycle.

Act Finally, in the fourth step ('act'), the new state of the control unit and the new actions are broadcast to the respective modules in parallel, based on the result of the transition function.

After the modules perform their new operations, a new bundle of data is gathered to refresh the dashboard: new sensations and their qualities from the sensory modules, and new reports from the motor modules. Then the entire operational cycle is repeated.

The *Sense-Analyze-Compute-Act* cycle concept as defined here resembles that of the *Monitor-Analyze-Plan-Execute-(over-shared-)Knowledge* (MAPE-K) loop seen in the design of self-adaptive *autonomic systems* [8]. The schema of a cognitive cyber-physical system is depicted in Fig. 1. A formal description can be given in a framework like suggested in [1] or in the automata-theoretic framework described in [12].

Definition 1. *A cognitive cyber-physical system is called* complete *if and only if its transition function is defined for all combinations of its inputs.*

A complete cognitive cyber-physical system can, in principle, react differently to different inputs in response to changes in its input parameters.

3 Self-controlled Cognitive Cyber-physical Systems and Minimal Machine Consciousness

3.1 Minimal Machine Consciousness

Cognitive cyber-physical systems can be adequately self-controlled only when a full 'picture' of itself and its embedding can be derived (implicitly) based on the information from their sensory and motor units and on the potential actions they can initiate at a given moment.

Considering the *Sense-Analyze-Compute-Act* paradigm, it makes sense to distinguish four corresponding 'dimensions' that together serve as prerequisites for adequate self-control. This leads to the following 'self-$\star$' properties which one might consider for a cognitive cyber-physical system C:

- *self-knowledge*: C has complete knowledge of its current cognitive state as well as of the data produced by all its sensors (the percepts and their qualities) and motor units (the reports from all of them).
- *self-monitoring*: C is completely informed about the performance and status of its sensory and motor units over time and of its embedding in its environment as it is.
- *self-awareness* (or *self-reflection*): C behaves in a way that unambiguously reflects, resp. is determined by, its current cognitive state and the information gained by its self-knowledge and self-monitoring abilities, and that is 'aware' of the internal and external changes that it causes.
- *self-informing*: C globally broadcasts its cognitive state, to all modules of the system and whenever changes of state occur.

Definition 2. *A cognitive cyber-physical system C is called* minimal machine conscious *(MMC) if and only if it is self-monitoring, self-knowledgeable, self-aware, and self-informing.*

There are several reasons for using the term 'minimal machine consciousness' for the collective properties we distinguished. A major reason is that, together, they seem to represent the minimal requirements for a system to respond adequately under all circumstances. Furthermore, a cognitive cyber-physical system was defined as a finite-state system without further resources. Thus, the 'active' memory available to realize any sort of 'conscious behavior' is only assumed to be finite, i.e., 'minimal' when compared to intelligent systems with (potentially) unbounded active memory.

The four principles of self-control are necessarily *informal*. We envision that for any class of cyber-physical systems they are concretized, to the extent that they provide precise requirements for the system designers and are verified for the systems that are claimed to satisfy them.

Example 2. Several disasters of airplanes and space shuttles have been caused by the lack of self-knowledge and self-monitoring qualities, and the absence of cooperation among the modules of the flight-control system. For example, in 1986, the space shuttle Challenger exploded due to an unspotted malfunction of the spacecraft's rubber seals. No one on board survived. The recent crash of Lion Air Flight 610 was caused by a malfunctioning of the flight-control system of a Boeing 737 MAX 8 that should not have happened if it had been a minimal machine conscious system.

Example 3. Cognitive cyber-physical systems that are MMC are not necessarily restricted to having finite memory only. For example, note that a Turing machine can be seen as a cognitive cyber-physical system in which a finite-state control governs a finite set of sensory and motor units, namely the respective read/write heads on its worktapes. Operating over an input stream, the system is seen to be minimal machine conscious. Nevertheless, the work-tapes give it a potentially unbounded memory.

3.2 Self-controlled Cognitive Cyber-physical Systems are MMC

The four principles that define minimal machine consciousness (self-knowledge, self-monitoring, self-awareness and self-informing) correspond precisely to the properties that are required for full and adequate self-control in the various phases of the operational cycle of a cyber-physical system. We formulate this as follows.

Proposition 1. *Self-controlled cognitive cyber-physical systems are necessarily minimal machine conscious.*

In the remainder of this section we expand on our arguments in support of the Proposition. Afterwards we discuss how close minimal machine consciousness comes to guaranteeing full and adequate self-control.

(a) Self-controlled cognitive cyber-physical systems have self-knowledge

The information needed for self-knowledge includes its current cognitive state and the information produced both by its sensory units (the percepts and their qualities) and its motor modules (the reports from all of them). In a cognitive cyber-physical system, this is the data maintained in the dashboard. It gives the system the possibility to report any information about its functioning, at any time.

(b) Self-controlled cognitive cyber-physical systems are self-monitoring

In the state it is in, the feedback from the sensory and motor modules as it is supplied by the feature of self-knowledge, makes it possible for the system to monitor itself. Namely, from these modules the system gets the data about its current working conditions, and based on this information it can either prolong its functioning without any further special actions or take steps that remedy or adjust its operation. All this confirms the machine's certainty, or errors, in its actions and enables the repair of its own mistakes [2].

(c) Self-controlled cognitive cyber-physical systems are self-aware

The current cognitive state and the information gained by its self-knowledge and self-monitoring abilities, enable a system to determine ('compute') whatever its appropriate next action would be, in the environment in which it operates. The property of self-awareness requires the fulfillment of three conditions:

– *the capacity of introspection*, i.e. the ability to reflect on one's own mental state (cf. [2]). General mechanisms of introspection seem to be beyond the ability of finite-state devices, as they may require unbounded memory. In the framework of finite-state devices, introspection can be modelled by a finite number of system states. For instance, 'interesting' past states can be stored in the current state, by using the standard automata theory technique of storing data in an automaton's state. In this way, one can even introduce dedicated states of the control unit, so-called *machine qualia states*, in which a system can remember important past events that still require its ongoing attention (cf. [12]). Machine qualia offer the system a mechanism for remembering certain 'subjective' cognitive states of the system that are bound to certain previous cognitive 'experiences' (states). For instance, a quale state in a mobile phone may keep a remembrance of a recent event when a text message was received. A driverless car can have a quale state regarding a shortage of gas. The qualia states stored in the system's global state can then be broadcast to the entire system as long as a circumstance invoking them persists.

– *the ability to recognize oneself as an individual object separate from the environment and other objects*. This will be implied by a proper selection of sender-receiver modules whose cooperation provides the required effect. There are several modalities of signals that can have a similar effect. For instance, receiving a specific olfactory (or chemosensory), electric, optical, acoustic or haptic return signal may indicate the presence of other instances of the system. Obviously, the absence of such return signals indicates that no similar systems are around. For a similar purpose, in advanced cyber-physical systems a vision system may be available.

– *awareness of changes in the outside world*. The feedback also allows the system to distinguish its actions as registered by its sensory modules from the similarly registered actions performed by other systems. That is to say, in the latter case, the reports from the motor modules do not match the sensations from the sensory modules. Self-awareness thus provides a cognitive system with a rudimentary machine concept of the *self*: the system has information on what goes on in the outside world, what its actions are and what their effects. This information is of the form 'here and now' – it is pertinent to the present position of the system in its environment and the present moment.

(d) Self-controlled cognitive cyber-physical systems are self-informing

By the very definition of cognitive cyber-physical systems, the new state of a system and the projected actions are 'broadcast' to all its modules, simultaneously and in parallel. This ensures a synchronization of the actions they need to be synchronized and gives the modules a certain minimal information (namely, that 'stored' in the current state) of what goes on in the entire system. Endowing machines with the possibility of self-informing allows their modules to share information and collaborate to address whatever impending problem (cf. [2]). For example, consider a modern car in which a fuel sensor reports a shortage of gas. If this information is globally available then the navigation system of the car can direct the driver to a nearest gas station [2].

4 A Manifesto On Minimal Machine Consciousness

We claim that minimal machine consciousness is a key criterion for all cognitive cyber-physical systems in practice, to provide them with the necessary abilities for smooth and safe operation. It has important consequences for the engineering of cyber-physical systems. We summarize this in the following assertion.

> *All cyber-physical systems operating in a given environment, with or without human aid, must be designed as minimal machine conscious cognitive systems.*

4.1 What Minimal Machine Consciousness Makes Possible

In Table 1 we list a variety of important abilities which cyber-physical systems should have and the mechanisms that facilitate them if the system is (cognitive and) minimal machine conscious. It shows what is made possible by the combined properties of the architecture and the four principle of minimal machine consciousness.

The feedback from the sensory and motor units brings straightforward benefits for improving system performance. Self-knowledge, self-monitoring, and self-awareness lead to improved decision making and increased detection capabilities. Self-informing enables the cooperation and synchronization of the system's modules. Altogether, minimal machine consciousness enables the system to detect and correct failures that can potentially prevent a possible crash or disaster, and at least diminish the number of false alarms, thus improving the trustability, reliability and safeness of the system under the changing conditions in its environment.

The detailed mechanisms of self-awareness lead to further potential benefits. For instance, self-awareness requires that the system must be able to distinguish its own

Ability	Mechanism
Improved decision making	Graded feedback from sensors and motor units
Increased detection capabilities	
Diminished number of false alarms	
Failure correction	
Damage registration	
Flexibility and improved reliability in varying situations	
Interception of adversarial physical actions	Additional sensors
Limited cognitive and calculatory tasks	Finite-state data processing
Attention mechanism	Suppressing disturbing inputs
Limited form of introspection	Cognitive states
Detection of patterns in ongoing processes	Introspection
Recognition of itself as an individual subject, separate from the environment and other systems	Cooperation of send-receive mechanisms
Communication	Send-receive mechanisms
Distinguishing one's own actions from the actions of other systems	Mismatch of motor actions with sensory observations
Reading the intentions of other similar systems (machine empathy)	Situating the system into the position of the other systems
Limited cognitive and calculatory tasks	Finite-state data processing
Subjective machine perception (machine qualia states)	Storing states in states

Table 1. Abilities of minimal machine conscious cyber-physical systems and the corresponding mechanisms that realize or facilitate them

movement from any other movement that it can observe in the environment. This property can be used, e.g., by a robotic arm system to intercept a motion (of an 'intruder) within reach of its arm. Another example is collision-free navigation. As an extreme case, a minimal machine conscious system can 'read the mind' of another, similar system by observing its input and by being aware that this is not its input. Namely, thanks to the fact that the observing and the observed systems are of the same construction, the observing system can infer the actions of the observed system.

4.2 Design Considerations

As presented, minimal machine consciousness becomes feasible once a cyber-physical system is, or can be, designed as a *cognitive* system. This follows from the close connection between the four principles of minimal machine consciousness and the necessary features of *self-control* during the consecutive phases of the operational cycle of the system. Minimal machine consciousness enables the system to operate awarely in its environment at any time.

We therefore contend that minimal machine consciousness should be one of the major *design objectives* of any cyber-physical system. Having this is mind, the following bold statement is at the heart of our 'manifesto'.

Claim 1 *Any cognitive cyber-physical system operating in a given environment, with or without human aid, can also be designed as a minimal machine conscious cyber-physical system.*

To see this, consider any cognitive cyber-physical system operating in a given environment that is not minimal machine conscious. This means that the system has knowledge of what behavior must be invoked, based on the inputs from its sensory modules, assuming problem-free operation of both its sensory and motor modules. It should thus be possible to 'redesign' (or, re-program) the system so as to make optimal use of this information and transform it into a system that conforms to the four self-control principles of minimal machine machine consciousness described above. We even *hypothesize* the following, stronger claim:

Claim 2 *All dedicated activities that can be consciously controlled by humans can also be controlled by minimal machine conscious cognitive cyber-physical systems.*

The background for this claim is that, if we take 'conscious control' to mean as much as 'possessing knowledge of how to behave to fulfill a certain task', then one must be close to knowing or discovering the dependencies between various components of behavior and the corresponding inputs from sensory and motor modules. If we accept the cognitive architecture as a *standard*, then the rules and necessary information to drive the operational cycle should be in reach.

The claims can be the starting point for further methodological, or software engineering considerations towards the realization of cyber-physical system as described, for example, in [6] and [7]. Including minimal machine consciousness as a concrete design objective calls for more orderliness and discipline in the design of the system, by insisting on the fulfillment of the four necessary conditions required for this type of 'consciousness'. The benefits are clear.

Example 4. It seems that, currently, no clear-cut example of a deliberately designed minimal machine conscious cyber-physical system exists. The closest example seems to be the modern smartphone. These phones usually have a 'dashboard' in the form of a *status bar* (e.g. along the top of the touchscreen). Here, the statuses of various important sensors, such as the quality of the wifi, mobile network, GPS, the bluetooth signal, battery power, etcetera, are depicted with the help of the respective icons. At the same time, the icons indicate the quality of the respective signals. This is, in fact, global information describing the current state of the device that is accessible to all modules of the phone, and a witness of the system's self-knowledge and self-monitoring. Last but not least, the system is quite self-aware – it can recognize the incoming calls, send and receive messages, establish the bluetooth connection, identify changes in its location (via GPS), etc. Interestingly, these abilities of smartphones are the result of incremental, technological evolution and the development of user requirements rather than of a purposeful effort to make the devices minimal machine conscious. This only confirms that the idea of minimal machine consciousness is a natural and useful concept that is worth to follow up and exploit as a design objective.

4.3 Minimal Collective Machine Consciousness

Given a number of different cyber-physical systems, there may be considerable potential in combining them into one 'composed system'. This happens, for example, when a complex task must be split over several cyber-physical systems, with each system dedicated to a well-identified subtask. This leads us to consider *networks* of cooperating cyber-physical systems.

If the 'nodes' are all cognitive cyber-physical systems, then one may turn the network into a cognitive cyber-physical 'meta-system' by adding a global finite-control that sees the nodes as sensory/motor units and combines their information into one global operational cycle (which need not be synchronized with the operational cycles of the nodes). This construction is especially interesting when the nodes are all minimal machine conscious.

Example 5. One may think of modern robots as cognitive cyber-physical meta-systems, with subsystems dedicated to specific tasks like vision, motion, sensing, and grasping. More generally, teams of robots, swarms of drones, nano-machines in a bloodstream, etc., also qualify.

To see what sort of additional machine consciousness this may lead to, consider an arbitrary cyber-physical meta-system D. The nature of the information D collects from its nodes may vary widely. It can be data from the specific subtasks of the nodes, statistics related to their activities, reports on the working conditions and cognitive states of the underlying cyber-physical systems, etcetera. Based on this, D can keep track of the part of the 'world' that is registered by its nodes. In particular, if the nodes are minimal machine conscious, D may collect their qualia states. As a result we get a networked cognitive system that is *minimal collective machine conscious.*

With our *manifesto* of cyber-physical system design in mind, one may reformulate both Claims 1 and 2 so as to hold for minimally collective machine conscious cyber-physical meta-systems, and even for *cyber-physical human systems* [11] as well.

4.4 Minimal Machine Consciousness and Industry 4.0

It is a serious challenge to design a specific minimal machine conscious cyber-physical system that can handle all possibilities, in the numerous situations that the system can face. This is well-recognized in the area of software engineering, for embedded systems and cyber-physical systems alike. However, the challenge must be met, as cyber-physical systems are crucial for all enterprises. This is notably expressed in the advanced concepts of smart manufacturing in *Industry 4.0* [5].

In Industry 4.0 it is foreseen that all production processes in factories are automated and computerized, making them flexible and efficient by the use of modern information and communication technologies and intelligent systems and services [10]. The processes will be connected and controlled by smart systems that manage entire production lines and make decisions of their own, in symbiosis with human operators. We refer to [9] for an overview of the technological challenges involved.

The core systems of Industry 4.0 can be recognized to be cyber-physical (human) systems. As we have argued in this manifesto, these systems must be designed so as to be 'cognitive' and, especially, minimal machine conscious. With the testability of the latter in mind [12], this seems certainly achievable when taken into account as a criterion from the outset. It may be the limit of what current hardware and software engineering methods can do.

Fortunately, a promising technology is emerging that enables both the design and the efficient testing of potentially minimal machine conscious cyber-physical systems: the technology of *digital twins.* In our case, a digital twin would be a digital replica of the physical part of a designated cyber-physical system. Using a digital twin of (a part of) the given or intended cyber-physical system, one can systematically test whether it will react properly to all external and internal conditions, provided the combinations and scenarios that the system's modules may face can be finitely enumerated. If a system passes such a test, we know that it is complete w.r.t. to all events that can be registered by the system (cf. Definition 1). Of course, if the testing cannot be exhaustive, the system can only be guaranteed as far as the scenarios went.

Digital twin research is a growing and flourishing scientific field (cf. [4]). Its application to the design and testing of minimal machine conscious systems can give further impetus to the research and development of this technology. It is generally accepted that digital twin technology is one of the key enablers of *Industry 4.0*. The concept of minimal machine consciousness has the potential to revolutionize this field further.

5 Conclusion

The main message of our manifesto is the following assertion.

> *All cyber-physical systems operating in a given environment, with or without human aid, must be designed as minimal machine conscious cognitive systems.*

Minimal machine consciousness is not a feature of only highly complex systems, and cannot be achieved by a mere software upgrade. Rather, it requires a different system architecture and a properly designed operational cycle that can deal with the graded feedbacks from all its sensory and motor units.

Minimal machine consciousness has a meaningful purpose, its benefits in industrial applications are substantial and can hardly be obtained differently. What is costly, however, is their development since maximal attention must be paid to their functionality under all possible conditions they can face, be they caused by software or hardware malfunction or unfortunate combinations of adversarial external factors.

When designing new cyber-physical systems, or innovating the existing ones, especially in which risks for human life are at stake, it is a matter of responsible design and ethics to make such systems minimal machine conscious.

References

1. Broy, M.: Engineering Cyber-Physical Systems: Challenges and Foundations. In: M. Aiguier *et al.* (Eds.), *Complex Systems Design & Management*, Springer-Verlag, 2013, Ch. 1, pp. 1-13
2. Dehaene, S., Lau, H., Kouider, S.: What is consciousness, and could machines have it? *Science*, 358: 6362 (2017) 486-492
3. Dennett, D.C.: What Can We Do? In: J. Brockman (Ed.), *Possible Minds: 25 Ways of Looking at AI*, Ch. 5, Penguin Press, 2019
4. ERCIM: Digital Twins: Special Theme. *ERCIM News* 115, October 2018, https://ercim-news.ercim.eu/en115
5. i-Scoop: *Industry 4.0: the fourth industrial revolution - guide to Industry 4.0*, March 2020, https://www.i-scoop.eu/industry-4-0/
6. Jackson, M.: Behaviours as Design Components of Cyber-Physical Systems. In: B. Meyer, M. Nordio (Eds.):, *Software Engineering*, International Summer Schools, LASER 2013-2014, Revised Tutorial Lectures, Lecture Notes in Computer Science, Vol. 8987, Springer, 2015, pp. 43-62
7. Jackson, M.: Behaviours and Model Fidelity in Cyber-Physical Systems. In: *Computability in Europe: Computing with Foresight and Industry* (CiE 2019), Special Session: History and Philosophy of Computing, Durham
8. Kephart, J.O., Chess, D.M.: The Vision of Autonomic Computing, *IEEE Computer*, January 2003, pp. 41-50
9. Lu, Y.: Industry 4.0: A survey on technologies, applications and open research issues, *Journal of Industrial Information Integration* 6 (2017) 1-10
10. Rüßmann, M. *et al.*: *Industry 4.0 - The Future of Productivity and Growth in Manufacturing Industries*, Report, Boston Consulting Group, April 2015
11. Sowe, S.K., Simon, E., Zettsu, K., de Vaulx, F.F., Bojanova, I.: Cyber-Physical Human Systems: Putting People in the Loop, *IT Professional*, 18:1 (2016) 10-13
12. Wiedermann, J., van Leeuwen, J.: Finite State Machines with Feedback: An Architecture Supporting Minimal Machine Consciousness. In: Manea F., *et al.* (Eds.) *Computability in Europe: Computing with Foresight and Industry* (CiE 2019), Lecture Notes in Computer Science, Vol. 11558, Springer, 2019, pp. 286-297