Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

ALEXANDER N. RAIKOV [iD]1 AND MASSIMILIANO PIRANI [iD]2, (Member, IEEE)

[1]Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 117997 Moscow, Russia
[2]Department of Information Engineering, Marche Polytechnic University, 60131 Ancona, Italy

Corresponding author: Massimiliano Pirani (m.pirani@univpm.it)

ABSTRACT The goal of the paper is to find means for the unification of human-machine duality in collective behavior of people and machines, by conciliating approaches that proceed in opposite directions. The first approach proceeds top-down from non-formalizable, cognitive, uncaused, and chaotic human consciousness towards purposeful and sustainable human-machine interaction. The second approach proceeds bottom-up from intelligent machines towards high-end computing and is based on formalizable models leveraging multi-agent architectures. The resulting work reviews the extent, the merging points, and the potential of hybrid artificial intelligence frameworks that accept the idea of strong artificial intelligence. These models concern the pairing of connectionist and cognitive architectures, conscious and unconscious actions, symbolic and conceptual realizations, emergent and brain-based computing, and automata and subjects. The special authors' convergent methodology is considered, which is based on the integration of inverse problem-solving on topological spaces, cognitive modelling, quantum field theory, category theory methods, and holonic approaches. It aims to a more purposeful and sustainable human-machine interaction in form of algorithms or requirements, rules of strategic conversations or network brainstorming, and cognitive semantics. The paper addresses the reduction of the impact of AI development on ethics violation. The findings delivered are used to provide perspectives on the shaping of societal, ethical, and normative aspects in the symbiosis between humans and machines. Implementations in real practice are represented.

INDEX TERMS Cognitive semantics, category theory, human-machine duality, hybrid artificial intelligence, holonic systems, stability in dynamical systems.

I. INTRODUCTION

The introduction of digital economy tools and AI (artificial intelligence) in society have been increasingly focusing discussions on the issues of replacing people with machines, distorting ethics, degradation of natural intelligence, malicious use of AI, and even a drop in the level of spirituality [1], [2]. The main ways in which AI can negatively distort ethics do not come from some futurological uncontrollable super-intelligence or ''killer drones'' [3], but from the risks associated with people's trust in the work results of AI [4]. The introduction of AI generates decision risks and errors in data, the degradation of human analytical abilities, the possibility of misperception and misunderstanding.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott [iD].

In this context, the problems of human-machine duality, and the ultimate role of artificial intelligence (AI) in human and machines interactions, are explored through a technological challenge that confronts two separate lines of research. Accordingly, the research objective of this paper is to examine, by means of some literature review, the twilight zone that appears when trying to conciliate two technological approaches that proceed in opposite directions. In addition, this work provides a research methodology proposal that tries to achieve a first technologically viable solution towards this research objective.

The first technological approach considered, is a nonformalizable cognitive and organizational one, aiming at ensuring the acceleration of decision-making processes. An example of these processes are strategic conversations [5], [6]. This approach proceeds top-down from non-formalizable

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE *Access*

human consciousness and unconsciousness elements towards purposeful and sustainable human-machine systems.

The second technological stance, and associated line of research, is based on formalizable methods aiming to the hybridization of symbolic and sub-symbolic approaches, as the well-known neural networks, machine learning, ontologies, extended description logic, and fuzzy information. The presented approach in many respects abides by the well-developed theories near to the situated cognition framework of [7], [8], ecological psychology [9], and new visions that consider some relational nature at the grounding of abstract cognitive-capable representations with recursive nature [10]. A prominent role for the convergence between humans and the artificial is played in a constructivist playground, in particular in the context of research on Artificial General Intelligence (AGI) [11], [12].

These new visions are here technologically and practically materialized by the RMAS (relational-model multi-agent system) architecture that proceeds bottom-up from tiny intelligent machines towards high-end means. The RMAS architecture has been conceived especially for autonomic industrial distributed systems [13], and is currently proposed as a new sustainable computational test framework [14], which rigorous formalization and full-fledged realization is under development [15].

The collision of two different worlds, the non-formalizable and the formalizable, is constantly met but goes mostly unnoticed in human being's reality. Studying one helps to understand the other, and vice versa. In this work, a methodology for such a study is offered. If the study is simultaneous, the phenomena under interest are taken as a whole. This leads to emergent levels of explanation that may arrive at a new hybrid kind of reality. For example, there are two different mechanisms by which infant, who cannot communicate, and adult, who has knowledge and intuition, might be related. Researching these two possible mechanisms together is an intriguing direction for future research on cognition [16].

In recent years, the volume of researches in the field of AI has grown rapidly, and a lot of literature and surveys have appeared. On one side, this has been largely due to the increase in computer power, to the acceleration of communications, and to the growth of Big Data. On the other side, AI development is connected with the high expectations that envision the possible realization of Strong AI and of General AI (AGI) and their specific numerous dimensions [17].

In order to head for practical and consistent applications, it is important to distinguish between Weak (AWI), Narrow AI (ANI), General AI (AGI), Hybrid AI (HAI), Strong AI, and Artificial Super Intelligence. Current state of AWI, as well as speculations on the future of Strong AI and Artificial Super Intelligence, have been drawn from science, culture, and philosophy [18]. HAI is an AI that is immersed into the Hybrid Reality, where human ceases to be an observer and

becomes a cognitive part of the system [19]. Wrongly AGI, Artificial Super Intelligence, and Strong AI are considered as synonyms [20].

It is also worth mentioning the results of recent standardization efforts [21]. These efforts clarify that, on the one hand, AWI and ANI systems can only process symbols or sub-symbols without any autonomous understanding of what is going on, in terms of human-attributable meaning. This goes along the well-known problem about the uncertain epistemology of AI algorithms, and the substantial difference between artificial communications and current acceptation of AI [22]. On the other hand, Strong AI systems are still able to process symbols but, at the same time, should also vaunt authentic understanding of what they do, namely their semantics [23]. ANI systems are able to solve specific tasks, possibly much better than humans would do. AGI systems are able to address multiple tasks, of different nature, with acceptable level of performance. ANI can solve single but very important tasks, such as pattern recognition, playing chess and Go, helping purchasing, weather forecasting, transport control, making medicine diagnoses, computer vision, language processing, combing through digital footprints, predicting when a crime will be likely to take place, and so on [17]. Currently, ANI uses artificial neural networks to bind together artificial intelligence algorithms with logical knowledge bases, in order to model and optimize vehicle navigation. Similar techniques may be used to perform optimization of RFID systems that consists of Redundant Antenna Elimination algorithm, Ring Probabilistic Logic Neural Networks, and evolutionary optimization technique, namely Genetic Algorithms [24].

Recent astonishing developments have been reached with the generalized few-shot learners [25], which among many can boast the impressive Generative Pre-trained Transformer (GPT-3) application [26], or successful growing capabilities in self-learning of neural networks application of the new generation [17].

At the same time, many new problems are connected with collective behavior of people and machines, and their sustainable symbiosis. This symbiosis, is being discussed with increasing intensity due to its implication on Strong AI and Hybrid AI developments. In this context, the solution to questions faced by the scientists encounters problems in of the impossibility of representing non-formalized procedures. This kind of problems concerns the collective unconscious, spirituality, emotions, thoughts, feelings, free will, etc.. Nevertheless, it is still evident that current research cannot embrace these aspects of human cognition. This is the mission of Strong AI and Hybrid AI, in particular for the parts that derive from logic and artificial neuron-inspired paradigms of AI [1].

Comprehensive models of cognitive processes are shifting towards spaces of non-logical (non-formalized and cognitive) semantics, much deeper and beyond the reach of logic descriptions and formulas.

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

In particular, non-formalizable aspects of Strong AI have to be taken into deep consideration in order to solve human-machine duality and symbiosis.

As it is claimed in this work, it is increasingly more important to hoist a bridge between the banks of the formalizable and the non-formalizable, over the abyss of meaning. This requires exploring new ways in which transdisciplinary means of mathematics can help materializing and studying this encounter, for example the exploration of methods inspired by quantum mathematics for the generation of new insights and results in AI [27].

According to the aforementioned purposes and claims, this paper starts with an overview of cognitive architectures and of current trends in AI developments. By means of a brief survey in section II, a focus on the statement of central problems in current and future AI will be made. On this basis, in section III a position will be made about a new methodology that has the ambition to expose a convergent nature and a mechanism for the harmonization of human-machine duality and the removal of contradictions and distances between bottom-up and top-down technological directions. Some examples of the application of convergent methodologies and implementations are proposed and shown in section IV, along with a summary discussion. Conclusions are provided in section V.

II. OVERVIEW, RELATED WORK, AND PROBLEM STATEMENT

In this section, the foundations of urgent problems in AI will be laid, which this work seeks to address, by means of the introduction of a convergent methodology that will be used to study the crossing points between human and artificial cognition. This foundational objective will be achieved through a critical literature review that confronts new visions with respect to state-of-the-art research and results in AI.

A business that needs to thrive in the market sometimes gives up to strictly complying with ethics codes. This distorts the ethics and the social responsibility that companies should ideally implement. An obstacle and deterrent to such a distortion of ethics can surely derive from the threat of loss of reputation in business [4]. However, existing mechanisms for manipulating public opinion, using AI, adversely mitigate these threats. In addition to business reputation, the opinion of civil society becomes an obstacle to the distortion of market ethics, which makes it possible to balance negative ethical trends. This state of affairs requires an appropriate coherence that is achievable through self-organization of the subjects of society that can consistently form the public opinion. Methods and approaches of network democracy can support self-organization by taking into account the capabilities of AI [28].

Many researchers all over the world are expecting or forecasting that the accelerating pace in the development of smart machines will soon — till the middle or the end of this century — outrun all the human capabilities. The singularity, awesome bifurcation and historical change point, will eventually arrive, and it is impossible to predict exactly now how the human future might be following after this point [3]. Nobody knows how the Strong AI will affect the life and destiny of human beings; someday humans might not be able to keep up with AI(s) anymore, and cognitive intelligence could be not the most important human trait in the end [29].

Description logic and symbolic frameworks, with their many variants, can provide mathematical methods and tools for representation and reasoning, even in the case of fuzzy information [30], explicit, implicit, or speculative knowledge [31], or probabilistic and abductive reasoning [32]. For example, [33] addresses an example of complex decision-making using ontology and knowledge-based approach for creating safe building evacuation design. Although there is a clear lack of automation in the modelling of knowledge, the system still can include several developments of ontologies and rules representing design knowledge, captured from the evacuation field. Nonetheless, in such playgrounds, there is a growing need of new research methods towards technologies that can automatically and dynamically exploit real-time information in order to detect unconventional ways outs in emergencies, and to arrange effective rescue operations [34]. Constructing plans ''on-the-fly'' for the unexpected, in presence of irrational agents, is increasingly under focus, and symbolic methods might not be always a solution [35].

The challenge ahead is the imitation of the fast-and-frugal, though well-informed, decision-making that constitutes the essence of relevant effective ways in which the humans act [36].

A major problem in current AI is a misleading overlap between causality and correlation concepts. According to well-known work [37], a mark of causality is a 100% correlation between cause and effect. But such correlations are not always due to causality. For example, a good correlation between the number of storks arriving in the spring every year and the number of human babies born does not bring causal evidence that storks bring babies. Traditional ANI cannot transform a correlation into causation, and cannot ensure high-level of transparency or trust. This problem has been well described by [38], who is the father of the powerful and widespread Bayesian networks in AI.

Recent research is investigating the potentialities of quantum probability for causal reasoning, shows that a quantum probability model has the unifying power of encompassing many heuristic strategies and associative thinking that, in the usual case of classical Bayesian and hierarchical models, constitute a violation [39]. This reinforces and confirms the role of quantum semantics methods focused in the methodology here proposed, yet considered valuable by some recent literature [27].

An open problem for decision-making, is the understandability of learned opaque models. Setzu *et al.* [40] propose an effective symbolic framework for black-box explainability. Their future work aims to consider non-logic explanations, and to propose a research direction that considers causal

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE *Access*

explanations based on data observed by appropriately querying a black box.

Being causality and black-box querying an essential part in the bottom-up approach, which will be presented below, a similar research direction is considered promising also in this work. An intelligent agent will need to construct a causal model of itself and of the environment in which she/he/it enacts, though sometimes most parts of this model are non-formalizable.

Beyond these promising practical applications, still some fancy speculations endure about future symbiosis between digital technologies and humans, brought about by a vision of full-fledged Strong AI. Creating such a technology requires complete understanding of the foundations of human cognition and thinking mechanisms. It is likely that such kind of singularity point will emerge from some uncontrolled and self-amplificating emerging capability, starting from basic seeds of knowledge [41]–[43].

About three hundreds cognitive architectures have been created and studied [44]. The core human cognitive abilities, such as perception, attention, learning, reasoning, and so on are found treated. But there are still only a small number of cognitive architectures devoted to AGI. Among them, well-known are the Soar and the Act-r systems, NARS [45], LIDA [46], SiMA [47], [48], and so on. Many works are devoted to the study of separate aspects of cognition. For example, the ARCADIA [49] takes into account human attention, and CELTS [50] emotions. The majority of mentioned systems are formalized and rely on logic methods in their foundation. Logic frames, production rules, and logical inferences are based on symbolic means but are often hybridized with sub-symbolic and neural networks structures.

Nonetheless, the most complex and unsolvable problem for current AI are uncaused effects in human consciousness. A human can make appropriate but uncaused decisions [51]. A lot of phenomena such as cosmic vacuum, chaotic dynamics of thoughts, collective unconscious, natural neural interactions, particle-wave duality, cosmic strings, quantum non-locality, dark matter and energy, escape causal explanations. Examples of non-classical approaches try to represent uncaused phenomena by topologically-related mechanisms of projections and mappings. These approaches can describe different phenomena, in which classical (deductive and inductive) logic and non-classical (non-monotonic, intuitive, and abductive) logics do not hold anymore [52].

Considering such approaches, authors as [53] suggested that the mappings of objects or events between different n-spaces, representing different levels in hierarchy of ordering of the objects or events, change the traditional understanding of cause-effect relationships. The fruitfulness of these suggestions was demonstrated clearly on some dynamical systems, which must necessarily exhibit at least a critical point of singularity — for example topological cellular duplication, and brain fluctuations' properties.

One of the well-known instances that demonstrate the possibility of uncaused events is the well-known entanglement effect [54] and the Bell's test against the local realism [55]. They required spatially distributed entanglement, fast detection and unpredictable measurement settings. Recently, the advances in information technology have made possible an experimental proof of quantum phenomena. For example, the unpredictability of human "free will" has been confirmed [56], creating the ground for constructivist approaches that take into account the role of agencies of subjects in scientific observations of hybrid reality phenomena.

The article of Lake *et al.* [41] substantiates that current AI systems differ from human intelligence in crucial ways. The authors show that modern AI can help recognizing the connection between events or objects, but cannot be used to effectively build causal models yet, which only humans can easily do. The difference between recognition through prediction and causal model building by means of explanation is the central issue for the understanding of human intelligence.

Some researchers have criticized the scientific effort of creating systems of AI that try to be identical to human (natural) intelligence. One of the most prominent work in this sense was due to [57]. This author argued decades ago that computers, which do not have a human body and no cultural history, will never be able to acquire natural intelligence. Author's main argument is that computer cannot articulate human knowledge because it is partly tacit and latent.

Currently, the most common idea in creating Strong AI and AGI lies in neuromorphic paradigms. Huang [58] proposed the use of the method of "reverse engineering" — to literally copy the neural structure of the brain, with the hope that this copy will reproduce its functions through emergence. It can be noticed that this approach continues an evolutionary chain of developments in AI — symbolism, connectionism, behaviorism, and statisticalism — by creating on the same logical-formalized basis a new version of ANI: intuitionalism.

This is reinforced by authors like Fjelland [20], to whom Strong AI and AGI have not come so much closer to their goal, neither they will in the future. Strong AI and AGI cannot be realized as human-like intelligence because computers do not belong to a culture, and do not act in the world. Studies like the one conducted in [59], confirm that abstract concepts are grounded in interoceptive experience, which is a pivotal requirement in reasoning and acting for a living entity. This line of research was also characterizing the cognitive models of Maturana and Varela [60], an alternative to the ever-dominating interpretation of cognition as information representation and processing system. This worldview among scientists and philosophers has been growing into some nowadays prominent derivations like the concept of 4E cognition – embodied, embedded, enacted, and extended cognition [61].

In the human case, every individual structure has been created during long time and has been exposed to long-distant environmental factors (phylogeny). Thus, the difficulty of building human-level AI has to explore deep problems, far beyond a mere computational modelling and imitation of the

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

behavior of neurons and their connections. Such problems have also to be considered in light of some frameworks of interpretation of computational models that go beyond current digital technology [62]. In this view, the focus is not to identify how representations are constructed in brains, but how a physics-driven intention-building system works. This kind of vision derives, for example, from the work of Freeman [63], which maintained that there are no representations in brains, only meanings. Representations in brains are traded for a self-organizing mechanism of intentionality, which consists of a nonlinear dynamical model of brain's function centered in the construction of meaning rather than in some kind of representation. In this vision, consciousness is the dynamical process in which meanings are continually under construction in a chaotic trajectory through brain state space, and awareness is understood as the subjective experience of the momentary focus of the activity that constitutes meanings. Nevertheless, the paper of Freeman [63] still does not designedly address the ''hard problem'' of consciousness, as posed by Chalmers [64] who classifies the construction of meanings still among the ''easy problems.''

Major outlooks about the future of AI seem quite far-fetched, because they are based on the assumption that the human brain is a finite construction, and thus that its basic behavior can ultimately be known and predicted.

Nonetheless, human intelligence cannot be represented in finite or clear logical ways; the human thinking is not only performed locally by the brain or even by the body. Every neuron of the brain consists of a huge number of particles, and all the particles randomly communicate with particles of distant surroundings [54], [55]. Perhaps, such communications underlie the phenomena of unconscious and insights. Now it is completely evident that the human intellectual mechanisms is many orders of magnitude more complex than that of the best computer available. New frontiers are currently realistically opening in quantum and neuromorphic computing. Some approaches are considering using quantum technologies that possibly will provide penetration into a simulated object, for example, using the quantum entanglement effect [65], [66]. This will allow coming closer to solving the problem of creating an atomic twin of an object, brain, and person.

Unconscious and insights — as well as the inner speech — capacity of humans are the subject of numerous studies [67], [68]. For example, the Eureka effect manifests itself and proceeds usually in through the following steps (the Klondike case): a long search, where breakthrough thinking may require dedicating years to find a small cue on the decision (a wilderness trap); a little apparent progress, where the researcher has to waste a lot of time, and cue on the decision after little or no progress (plateau traps); a precipitating event, when some circumstances' cues give the chance of decision (creating a canyon trap); a cognitive snap happens in the moment when a breakthrough idea rapidly appears, and not much time separates this precipitating event from solution (an oasis trap); eventually, a transformation of the
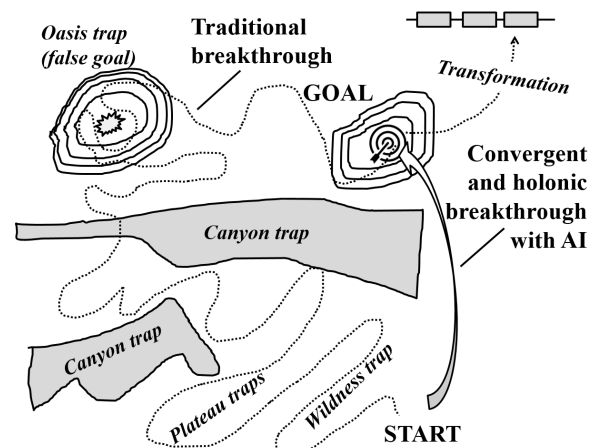

FIGURE 1. The Klondike's breakthrough and its accelerating effect.

mental or physical world proceeds in a generative way, whilst creating details of decisions and making. The diagram of this accelerating process is illustrated in Figure 1.

According to this phenomenology, the objective of current research is the study and development of means for accelerating the aforementioned breakthroughs, by using jointly two initially separated methodologies that respectively deal with convergent [28] and holonic approaches [69]. This research objective will become more understandable and clearer by means of some deeper elaboration of the concepts, starting from the next section.

The path towards Strong AI is mostly, and pragmatically, associated with the idea of hybrid AI, ensuring the immersion of man and machine into a single common space. The search for such a man-machine symbiosis is ongoing. Scientists are trying to remove the limitation of the discreteness of the presentation of data, information, and knowledge. Many problem areas in the future of AI are likely to be characterized by the following features and requirements:

- Retrospective knowledge (in symbolic forms) is insufficient as model of cognition, as well as connectionist approaches as neural networks.
- It is necessary to take into account the uncaused and chaotic cognitive semantics of AI model.
- The states of the AI model can be changed in an unpredictable jump-like (quantized) way.
- The behavior of the model is affected by objects which location and function are unknown (quantum non-locality).
- The AI model has to completely replace the managed object, and become its twin at the subatomic level.

The special convergent approach, which ensures purposefulness and sustainability of collective decision-making in hybrid society of humans and artificial agents, was designed to meet these requirements [6], [28], [70]. This approach is the ground for the convergent hybrid AI resonator developed hereafter. It goes towards the introduction of higher-order cybernetics lenses over the AI problem. This

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE Access

vision is supported by Dobbe *et al.* [71], where authors highlight that today's AI systems require a new lens that can see beyond technical practices, identifying points of overlap between design decisions and major socio-technical challenges, in particular with the need to become critical about certain formal assumptions behind intelligence. In addition, Dobbe *et al.* [71] claim that there is a need for the specification of requisite feedback modalities, in second-order cybernetics context, in order for the system to achieve appropriate stability.

In the present work, the category theory will be one key in order to create such kind of self-stabilizing circularity of the hybrid AI process. A Hybrid AI can help to find good decisions in the process of achieving an unclear goal, or speeding up the multi-group self-developing networked democracy decision-making processes [28]. In particular, the structuring of information in a special convergent way helps in reaching a collective strategic consensus between members of citizens' meetings, faster than usual.

Cybernetics, as here proposed, comes in aid when explainability is at stakes, in particular in the contact layer between the human and the artificial. Dazeley *et al.* [72] provide a bottom-up constructivist model for explaining AI agent's behaviour. A behaviour in which low-level "narrow" explanations of how individual decisions are taken are extended reflectively to the context, trying to provide insights into an agent's beliefs and motivations in relation to other (human, animal or AI) agents' intentions, interpretation of external cultural expectations, or processes used to generate its own explanation.

In section III-B, it will be envisioned how the cybernetics of second and third order admit that meanings can be constructed through eigenforms, comprising the observer's active role. This gives a new research perspective beyond state of the art in explanations and interpretations of the contact points and possible synergies between symbolic and sub-symbolic systems in AI.

III. CONVERGENT COGNITIVE AI RESONATOR

The convergent decision-making approach can be useful for speeding up decision-making in a hybrid society of human and artificial agents. The structuring of information is performed in a special way during interactions between humans and artificial agents. It can help human agents, in multi-level membership of groups, to reach a collective strategic consensus and generate ideas. The convergent technology is based on assembling methods of inverse problem solving on topological spaces, cognitive modelling, Big Data analysis, genetic algorithms, and holonic and self-adaptive holarchies approaches [6], [70], [73].

At the core of the convergent decision-making resides a deeper understanding of the methods that model, and then implement, the power of Strong AI.

In the following subsections, authors propose a methodology that is based on the convergence of two human-machine approaches proceeding in opposite directions. This method-

ology is expected to provide more insights and strengths at the meeting point of the two sub-methodologies.

Metaphorically, the intent is to obtain new information and insights by colliding, or reacting, two complementary elements, and trigger a resonance of the two sub-methodologies in the convergent cognitive AI resonator. Furthermore, it must not be forgotten that convergence is indeed a fundamental normative principle that, in the narrow sense, is necessary to guarantee the successful functioning of cognitive technology and, in the broad-sense, it is necessary to empower individuals and provide ultimate control and protection against malevolent applications of cognitive technology [74].

This occurs at the limit between AWI/ANI/AGI and Strong AI. Strong AI is expected to emerge from this resonance in a way that should be explainable for human beings. The emergence phenomenon should be kept under control. Lack of explainability will impact negatively on responsible use of AI in society, and definitely hinder the reproducibility and the scientific acceptance aspects of its phenomena [75].

The directions of the two above-mentioned human-machine approaches are attributed conventionally. As human is considered the top of natural intelligence in the common world of affairs, the first approach is deemed as top-down, as the intelligence is processed starting from organized and purposeful (but non-formalizable and uncaused) structures of discourse pertaining to the intelligent human beings. On the other side, the bottom-up direction processes the formalizable intelligence phenomenon, starting from structures of discourses arising from evidently less purposeful agents, like the artificial ones. In recent experiments, the cybernetics of the bottom-up and top-down in hierarchical goal-driven states have revealed an important role in human psychology. For example, in [76], bottom-up processing is the cognitive processing of sensory information, and cognitive processing capacity is automatically allocated to salient stimuli, while top-down processing is a more deliberate allocation of higher-order cognitive processing.

In the present research, the bottom-up paradigm is extended to consider the current state of the artificial technology as only a primary level of the cognition phenomenon. It is done with the aim to understand the mechanisms that will let it grow potentially up to higher orders of cognitive processing. Moreover, the encounter of top-down and bottom-up approaches in a hybrid land is under keen exploration when morality and ethics of AI are at stakes, although most of the studies until now remained at a conceptual and normative level [77]. As observed in [78], "the top-down approaches emphasize the importance of explicit ethical concerns that arise from outside of the entity, while the bottom-up approaches are directed more at the cultivation of implicit values that arise from within the entity."

The exploration of the hybrid dimension of the problem is thus promising to provide insights and some satisfaction to the need and urgency to regulate AI, as its complexity and impacts on social standards are far to be completely defined and then controlled [79].

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

A. THE TOP-DOWN APPROACH

The discussion of problems happens at the top level and in a conceptual realm. There participants operate with concepts and their connections; processes are non-formalizable and uncaused. These concepts do not fit into the rigid framework of metric spaces. The participants can be heads of government bodies, top management of companies, and so on. They may constitute collectives or groups in form of crowds, in which collective consciousness and unconsciousness affect decision making [28]. Often participants do not express plainly what they think or feel. They have their own interests that they want to promote. Representing these thought processes requires mostly abstract spaces rather than logical, fuzzy logical, or metric ones. These processes create a downward flow of conceptual non-formalizable information that must be understood and perceived at the lower levels of management.

Upper and lower levels of management use different languages and concepts. There is no well-established way to build a formalized chain of signal conversion between these control layers. The AI models have to help making a bridge between the upper and the lower levels of management. The models that are used in the decision-making discussions have two kinds of semantics, one of which obstructs the acceleration in getting agreement.

The semantics of any AI model can have a denotative and cognitive meaning [6]. The former means the mapping of the logical model to real things, objects, relations, terms, Big Data, and others, which constitute the volumes of the model's components. The latter is the subject matter of thoughts feelings, and emotions.

It can be noted that Big Data consists of signs. Thus, the AI model on this basis is denotative. The mental (thoughts, unconsciousness, etc.) aspects of AI semantics remain uncovered by Big Data although the numerical traces of them are captured in some way. Denotative semantics has a deficit of completeness, since mental processes are carried out beyond the bounds of reasoning. The replenishment of this deficit could be accomplished with cognitive semantics interpretations of AI models. However, the whole phenomenon cannot be recorded using signs and symbols. It can only be coped with in an indirect way, for example, with using the convergent approach that creates necessary and sufficient conditions for purposeful and sustainable decision making [6], [70].

The dynamics of the decision-making groups combine the analytically divergent and the convergent synthesis of processes of getting participants' consensus. The divergent tendencies and non-metric (non-quantity) descriptions of the situational problems restrict any attempt from participants to speed up the process of decision making. Such a process implies the coordination of interests, goals, tasks, actions, resources. These reflect free will, emotions, desires, thoughts, and intentions of participants.

In order to accelerate the collective decision-making process, the problem has to be split into many parts (divergent analysis), and then all the results have to be collected in a holistic solution (convergent synthesis). Non-formalizable and uncaused cognitive semantics must be taken into account in this collective decision-making process.

Current researches, concerning mentality, rely on various approaches in order to study the brain and thinking by using formalizable tools. Nonetheless, not many of these researches take into consideration atomic mechanisms of individual or collective brain behaviour. The number of neurons in human brain is in the order of 10^{11} and the number of atoms about 10^{26}. Atoms are under the influence of fields: electromagnetic, gravitational, strong, and weak. Electromagnetic waves are subject to distortion, diffraction, and interference. The behaviour of particles is fluctuating. The behaviour of atoms is influenced by the effect of an observer, which can be another system or a set of sensors [80]. The states of atoms can be entangled with atoms from the outside [65].

The atomic elements that provide a thought process seem to form clusters, which are holistic but at the same time informal objects. With this assumption, it is possible to identify a phenomenon of thought in a space of physical and mathematical interpretation [81].

Typically, the well-known low-frequency signals that human brain sends outwards look like communication signals. These signals closely resemble to speech by which a person transmits her/his thoughts in words to another person. These signals are not thoughts; feelings, and unconsciousness manifest themselves with thoughts that can be represented in the form of a physical field described using the methods of classical electrodynamics, quantum field theory, quantum optics, and even the theory of relativity. The wave aspect of thinking can be considered as an acoustic, electromagnetic, and quantum-relativistic resonator. To describe such a phenomenon, one can turn to the fundamental principles of field theory, concerning the wave nature of signal propagation. Such a basis is constituted by the D'Alembert and Helmholtz equations, the Green's function, the formalisms of electrodynamics and optics, both in classical and in quantum form. These means can be seen as a possibility of expanding the space of cognitive semantics interpretation. The heterogeneity of the medium is allowed, which in practice implies some dependence of the signal propagation speed on various factors, including the coordinates and influence of the observer (participants). The behavior of the field can be quantized, it can have an analog (continuous) and discrete character.

In the quantum field context of the representation of the cognitive semantics, particular interest is in the effect of the collapse of quantum states, as well as the effect of the entanglement, which reflects the instantaneous relationship of the states of quantum particles over large distances [54].

The collapse of the quantum state of quantum particles is reminiscent of the "Eureka" effect [67] when a person finds a solution to a problem instantly, after long reflection and due to an unexpected external shock. Actually, in the case of quantum physics, it is believed that information from one

participant in the events is not transmitted to another; just both participants at the same moment in time will see the same quantum state of the system from some set of their superposition.

In this view, the central element of the collective thought's (cognitive) space is based on the concept of an "event." It is the physical phenomenon distributed in some mentally limited space and time. The thought or event is limited. It has certain conditional boundaries. This event with its coordinates and time aspects depends on an observer, who wants to describe the event, but cannot represent it in a direct and logical way. This view allows placing a symbolic AI model with its denotative and cognitive semantics in a single space, using unified operators to implement transformations of events of various natures. In this case, events can be attributed to phenomena, both described by logic, the laws of quantum electrodynamics, and the special theory of relativity. The category theory and topology theory are applied to ensure the purposefulness of decision making, with taking into account both the cognitive and denotative semantics of AI models [82], [83]. The category and topology theories help to study the properties of the relations of objects without their detailed logical structures. In this case, cognitive semantics corresponds to a certain mathematics category, for which the following axioms hold:

- the number of elements in the system is infinite, and the maps of the objects are the maps with the closed graph;
- there is a non-empty finite sub-covering of the set of elements (*bicompactness*);
- any point can be associated with some neighbourhood (every open set contains this point) such that for each two points there are always disjoint neighbourhoods (*Hausdorffness*).

In addition, the following axioms show some requirements of the convergent approach:

- In order to get testable and verifiable AI models, the axioms need to be anchored in the context of solving problem, for which the fuzzy approach, for example, can be used.
- In hybrid AI system the cognitive semantics should be built by humans in a convergent structural way [28], [70]; in this case it can be verified and validated in the decision-making processes.
- There is the need to establish ontological links between the non-formalizable space of cognitive semantics and formalizable space of denotative semantics; the former is a mental process organized taking into account the listed axioms, and the latter proceeds bottom-up through multi-agent systems.

These axioms and requirements help to make the decision-making processes with using hybrid AI in a more purposeful and sustainable way, and help to provide a smoother consistency of the information flows encountered when combining the *top-down* and *bottom-up* approaches in the control.

B. BOTTOM-UP APPROACH

This section indicates a path that should enable simple-enough machines to cause and, at the same time, to capture the phenomenon of emergence in AGI context. With pragmatically avoiding here the philosophical *hard problem* of consciousness, we limit ourselves to a functional definition of intelligence that induces cognition, following [84]: AI can be functionally defined as the automation of cognition (machine cognition) to develop skills and competencies to perform (complex) tasks.

Starting with a core set of seminal constituents, the bottom-up approach aims at harnessing the emergence of intelligence phenomena, from simple machines up to increasingly complex cyber-physical systems of systems, and searches for invariant mechanisms that would bring convergence into natural-like intelligence, spanning through the whole range of living beings, from the bottom up.

Palacios *et al.* [85] proved that recursive and hierarchical self-organisation of Markov blankets can explain the self-evidencing, autopoietic behaviour of biological systems. They explain that such a common hierarchical self-organisation is a recursive process that can repeat itself at higher levels of description, with the absence of a privileged point of view — the dynamics at every level play the role of macroscopic states at the level below, and the role of microscopic states at the level above. Moreover, it is plausible that control relationships within autopoietic mechanisms give rise to dimensions of organization that are missed by existing accounts of mechanistic and causal explanations: organisms must control (also build and maintain) themselves and procure their own energy throughout their hierarchical structures, which is the biological autonomy problem [86].

Similar kind of natural recursive structuring of stable organized systems is well-known (Simon, 1969). This vision inspired another widespread concept, that of the *holon*.

As we will see in the following, the *holon* can be one of the names of previously-mentioned recursive connection entities between hierarchy levels of an organized system of systems [85]. Another example where we can spot the presence of a *holon*, in our own view and interpretation, is an olfactory quale resulting (emerging) from information cycled through a hierarchy of neural networks that reach a resonant state, as treated by [87]. In this case, the holon is the interpreter, mediator, and glue between the physical and the informational realms (intended as abstractions).

The holon can have also the role of the entity that tries to fill the gaps in the creation of representation functions in abstractions during the modelling process, as foreseen by the Abstraction/representation theory of [88]. In the extensions to this theory provided by [89], such representational entities can be called agents; it results in the Agential AR (AAR) theory. We are prone to see these kinds of agents as holons.

Although representations are not in principle linked to a specific human capability, Szangolies [90] holds that qualia are at the basis of representational capacities and cannot be

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

realized with whatsoever computational means. This would doom holons, handling such qualia, to remain a human (or at least higher animal) entity. Szangolies [90] tries to avoid the infinite regress that would be required to achieve the representing map that models a mind process. As we will put forth soon after, it is exactly in this infinite regress that resides the interest in the methodology here exposed.

It is worth noting that what is indeed fundamental is the search of an automation of an explanant that would in principle allow such an infinite regress; although, allowing it to stop, finitely and pragmatically, at an approximate *satisficing* states of equilibrium between the computational representational agent and its environment.

1) WHAT IS A HOLON? THE AMPHIBIAN ACROSS COGNITIVE WORLDS

To authors' knowledge, the holon concept first appeared in [91], [92], where author defended that life must be hierachically organised, with certain principles or laws which define the meanings of 'hierarchic order'. This hierarchic structuring was in turn inspired by [93].

Any two levels of the hierarchy can be interpreted as wholes and parts. Any entity (abstract or natural) that simultaneously straddles two adjacent levels of a hierarchy must be featuring the so called *Janus effect*: each element of the hierarchy has a face downwards the sub-level that sees a self-contained whole of sub-assemblies, and a face upward the apex acting the role of the part. Thus, holon (from the Greek holos=whole, on=part) is the term coined by Koestler to account for such Janus-faced entities that feature at once ("according to the way you look at them") the whole and the part role.

These kinds of entities have inherent disposition for the occurrence of emergence phenomena. The holon is an "active" interpreter between two realities, between the macro and the micro level, between two formal systems, or even between two not representable but "live" objects in the state of affairs. The meaning of "active" and "live" will be more clearly stated in a while.

The use of the holon concept is quite common in MAS (multi-agent system) context where a holon is implemented as an agent that is an "autonomous" program running a piece of software [94], [95].

A most notable and practically used concept enabled by the holon is that of the *holarchy*. The holarchy is a temporary purposeful grouping of holons into a hierarchical relationship (or in general a directed acyclic graph relationship). The holarchy concept is particularly useful when used as a functional structure for the expression of autopoiesis and emergent self-organization in systems. In particular, the holarchy can be associated to a teleological behaviour that some system or part of it has to perform in order to reach organization towards a goal.

Holarchy can be defined and visualized, informally, as a temporary purposeful cohort of holons. Holarchies acquire meaning through the specification of their purpose, and as such holarchies are inherently endowed with a teleological mechanism. The whole holarchy is a purposeful holon, which in turn can be a part of even bigger wholes, recursively.

The holon in a holarchy straddles three formal systems: the language of the parent, his own model of reality, and the model of the children or of the controlled subsystem with its own behavioural rules. If this property is used at its limits (in mathematical sense) along with recursive and fractal holarchies, special forms of multi-strata modular holarchies, composed of sibling elements, can be arranged to control an infinity of reality levels [69], [96].

Each holarchy level, and its knowledge model, addresses a level of the reality (state of affairs) with a minimum viable approximation (the stable formal system of *knowing-and-being*), once suitable granularity of the epistemic mechanisms of the holonic agent are set.

The openness in the granularity and in the number of the levels of the holarchy constitutes its major strength. Indeed, a major weakness of all the formalizable languages, is that they to undergo incompleteness and inconsistency to which any strong enough formal system is ineluctably doomed to.

Some state-of-the-art examples are found in the field of description logic framework [32], [97], [98]. Nonetheless, in cognitive approaches, such as the ones in which description logic is used, the main problem is the operational closure of every symbolic system.

This suggests that a suitable technological epistemic invariants should be searched that allowed the dynamic switch across all the needed symbolic levels of holarchies. Piantadosi [10] provides a clear example of such kind of invariants and their capability to move across different semantic contexts. What is desirable, in order to render the switching between semantic contexts, is a mechanism for the swift passage across them. Such a switching must be triggered and occur as soon as unavoidable difficulties (formal paradoxes) are detected in local symbolic systems.

Thus, the idea here put forth is to weaken the symbolic domains straddled by holons, but let their holarchy be able to traverse quickly symbolic systems up and down in the corresponding holarchy levels, switching nimbly across meta-language levels.

In order to support this feature, it is important to focus on generative and constructivist aspects of knowledge, in view of some purposeful version of AGI. According to [41], the actual open problem is that "It is crucial that learning-to-learn occurs at multiple levels of the hierarchical generative process." In addition, while the symbolic approaches remain explainable in some sense, the connectivism inherent in most of the recent ANI (machine and deep learning, neuromorphic computing) loses this property.

Cognition is an active process, both in the case of implicit or symbolic versions of it. This has been acknowledged recently in neuroscience, as active inference is a constructive approach to understanding behavior [99]. The process of knowledge acquisition occurs with recursive computations, along with the active role of the observer for the creation of a

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE Access

stable reality from the observer-agent's point of view. Cognition can be theorized as the product of recursive computations of computations [100], [101].

The observer's problem is at the basis of the second-order cybernetics phenomenon [102]. Furthermore, in third-order cybernetics the model of social recursive interaction among agents can be modeled as a self-developing poly-subject reflexive-active environment [1], [103].

The core of the artificial cognition problem is to establish a framework for an artificial *Self* [11]. The artificial *Self* for AI should be provided by crossing the limit between physical and artificial mental realities. The entity that crosses the boundary continuously is the so-called eigenform [104], [105]. In a world of eigenforms, the observer and the observed are one inseparable entity, in a process that recursively gives rise to objects. An object is a symbolic (or also sub-symbolic here) entity, participating in a network of interactions, taking on its apparent solidity and stability from these interactions.

Kauffman [104] defines the eigenform as the amphibian that crosses the boundary of mental to the physical, but also as the entity that crosses any two levels of semantic realities. In this scenario, it is straightforward to identify a similar role for the holon as that of an amphibian, and let its implementation encompass that of an eigenform.

An entity (the eigenform) makes queries to the environment with senses and actuators, and receives replies as information and associated change of internal state: this is the act of *knowing*. When the queries are repeated, and the answers do not change, the entity is in a stable state of *being*. The received information has organized the entity itself with respect to its interaction with the environment. This organizing force is usually measured as a reduction of entropy.

If some answers do not steadily arrive, or are not as expected, the entity experiences "irritation" — as the irritation state is defined for cognitive agents by [106]. In response to the irritation the entity has to autonomously start a new knowing procedure in order to reach soon a different, new, stable state of being.

In the AI practice, this kind of entity can be implemented as an agent that realizes this "living" process of attainment of stability between knowing actions and being state. This stabilizing process will need the switching across different levels of realities belonging to the holarchy in which the holon is defined.

The process of knowing and being for such an agent is constructive and generative: "we change the world and the world change us" [104]. This vision is growingly supported today also in physics, for example [80], [107] and [56].

The action that generates agent's reality can be expressed mathematically: if X is the *being* and f is the act of knowing then $X = f(X)$. If f is applied indefinitely (though at established sampling events) then X is said the fixed point for the invariant operator f when $X = f(f(f(f(f(f(f(f(\dots))))))))$. X is a fixed point of f as a value that is mapped to itself by the function f – "An eigenform is a fixed point for a transformation" [104]. Moreover, f is "at a level where the

level and the metalevel are one." The operator f is both object and subject of the discourse. The set of eigenforms that remain stable with respect to f constitute the reality for the agent [104].

The agent is immersed in a *reflexive* domain, which is an environment that is influenced by the presence of the agent. The reflexive domain is an abstract description of a conversational domain in which cybernetics can occur; in full reflexivity, each participant is entirely determined by how he or she acts in the domain, and the domain is entirely determined by its participants. Moreover, a reflexive domain is a particular situation where *lambda* calculus applies; a reflexive domain is itself an eigenform and can be transcended to a new and larger domain with endless process – this property is guaranteed by the fixed point theorem of Church and Curry [108].

In the linear context (e.g. observables in quantum mechanics), the eigenform becomes an eigenstate associated to a linear combination of eigenvectors of finite or infinite dimensions. For an autonomous agent to be in control of the state of affairs, it (she/he/it) has to experience equilibrium between the *being* (i.e, acting effectively in the environment) and *knowing* (i.e. having a suitable model of the environment in order to use it for the control, or for self-organization, or autopoiesis). So agent's organization – consciousness in some sense [60] – can be a transformation (in eigenspace) of the function of querying (*knowing*) about *being*, like $K(\mathbf{b}) = \lambda \cdot \mathbf{b}$, where K is the act of *knowing*, by querying the reality (sensing, learning-by-doing, trying, etc.), and $\mathbf{b}$ is the formal vector of representation of *being*, having a contextualized local meaning.

In the relational quantum interpretation [80], [107], the K operator would be the measurement of a quantum phenomenon influenced by the inquiring observer. When eigenvalue ceases to exist or does not guarantee stability, this determines the triggering of systems' switch.

An interpreter, namely the holon, that could continuously move in and out of a certain symbolic domain can be materialized conveniently as a holarchy. Holons make possible some controlled transitions between the hierarchical levels associated to language and meaning domains, along with the emergence or reduction from one into another.

Up to this point the focus has been on description logic or symbolic realms. The rest of AI world has been kept out, for example the above-mentioned AI methodologies like connectionism, behaviorism, statisticism, and imitationalism [58], but also the more challenging and not formalizable ones pertaining to natural intelligence and expressed in section III-A.

Then there is the need to introduce some further tools as the category theory, the Relational Model, and lambda calculus, in order to set up an appropriate integrated framework for AI, as discussed in the next section.

2) TOOLS AND METHODS IN THE BOTTOM-UP APPROACH
General systems theory, category theory, its monads, and the lambda calculus allow to operate with objects that are not completely specified, unknown, implicit, or dynamically

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

"alive," as intended for the eigenforms previously treated. Under this perspective, the proposed toolbox for AI, particularly oriented to autonomic computation, should comprehend:

- *general systems theory* (GST), in order to describe (holonic) systems defined in general reflexive domains of eigenforms;
- *category theory*, in order to map symbolic systems, monads, and eigenforms into dynamic systems;
- *stability theory*, in order to end the recursions and assure the state of being for an entity (e.g. Lyapunov-based methods);
- a "universal" control technique, in order to pass from "irritation" states into stable states of being with an invariant mechanism across the holarchy;
- a multi-agent implementation of holons, in order to manage and couple symbolic processing, sub-symbolic or non-algorithmic processing, and lambda calculus, with a scalable and flexible granularity implementation, ranging from tiny embedded systems to high performance computing.

The last item is specifically treated in the next subsection, but a brief position can be made here for the first four items of the list.

Following [109], a first important position is that different systems can have very different methods of specification. Nevertheless, the bottom line in [109] is that, in general, a system S can be expressed in form of *relations*, as $S \subset X \times Y$, with X and Y being the input and the output objects respectively. This expression holds also in the case of incomplete information or when the system can only be described in terms of a set of verbal statements. Still, these verbal statements, by their linguistic function as statements, define the system as a relation.

The concept of dynamical system is brought about in order to take into account system's behaviour developments and evolution in time. This is necessary in order to establish relations between the values of system's objects at different times. With defining a general system as a relation on abstract sets, then over sets of abstract time functions we can define the general system's time. In this way, an association with a suitable dynamical system is possible in almost any case.

In order to enable more specific definitions of various types of systems, certain kinds of so-called auxiliary functions are introduced, for example state transitions. This can bring to a structured model in state space (typically affine nonlinear, or linear), namely $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t) + B(\mathbf{x}(t), t)\mathbf{u}(t)$. Thanks to category theory, a set of symbols constitutes the alphabet of the system's states and variables – e.g., the temporal axis for ordinary differential equation representations in typical control and systems engineering problems. Category theory gives clearer insights into why apparently similar systems sometimes behave differently, and why some apparently very different systems share common structures [27].

There are a number of important discrete-time processes, such as computation, theorem proving, symbol-manipulation processes, and the like, which can be represented by dynamical systems as well. By means of the Fundamental Diagonalization (Gödel) Theorem some classes of symbolic systems can be mapped into dynamical systems.

As a representative of a formal system, author shall use a representation defined as an ordered sextuple [109], and let $K = < E, S, T, R, P, \phi >$, where:

- E is a denumerable set and represents expressions;
- $S \subset E$ represents sentences;
- $T \subset S$ represents theorems of S;
- $R \subset S$ represents refutable sentences;
- $P \subset E$ are (unary) predicates;
- N denotes the set of integers.
- g is the Gödel (restricted) function $g : E \rightarrow N$; and $\phi : E \times N \rightarrow E$ such that $\phi(e, n) \in S$ whenever e is a predicate, $e \in P$.

According to [109], it is easy in principle to construct a general system for K by establishing the following correspondences:

- predicates P are inputs of the system;
- expressions E are the states;
- sentences S are outputs;
- the Gödel (restricted) function g is such that the state representation $\rho : E \times P \rightarrow S$ is $\rho(e, p) = \phi(p, g(e))$;
- the theorem set $T \subset S$ corresponds to an equilibrium set Y_0.

Having obtained a suitable and practically usable expression of symbolic system in form of dynamical system, category theory can now be used to handle functorial homomorphisms (back and forth) between the starting domain (descriptive state of affairs) and its dynamical (continuous or discrete) state-space representation. This is expressed in Figure 2. The objects in each category (i.e., the element of systems of a given type) are related by morphisms in respective categories. The categories of systems are then related by functors, the constructive functor and the forgetful functor, respectively. The constructive functor maps the systems with less structure into the systems with more structure.

A parallel can be easily made to the Encoding and Decoding mappings of AR theory of [88] that correspond to the forgetful and constructive functor respectively. An explicit link between AR theory and category theory was envisioned also in [110].

In this specific case, a constructive functor maps general formalizable systems into dynamical systems. The forgetful functor maps the systems into opposed direction [109].

A simple example can now be made in order to describe how the use of category theory and GST can be associated to the holonic concept and to the control and stability theory.

We consider an "intelligent" machine endowed with the capability of producing a stable eigenbehavior. This can be, for example, a thermostat with its automation implemented through a digital program. The semantics of this program
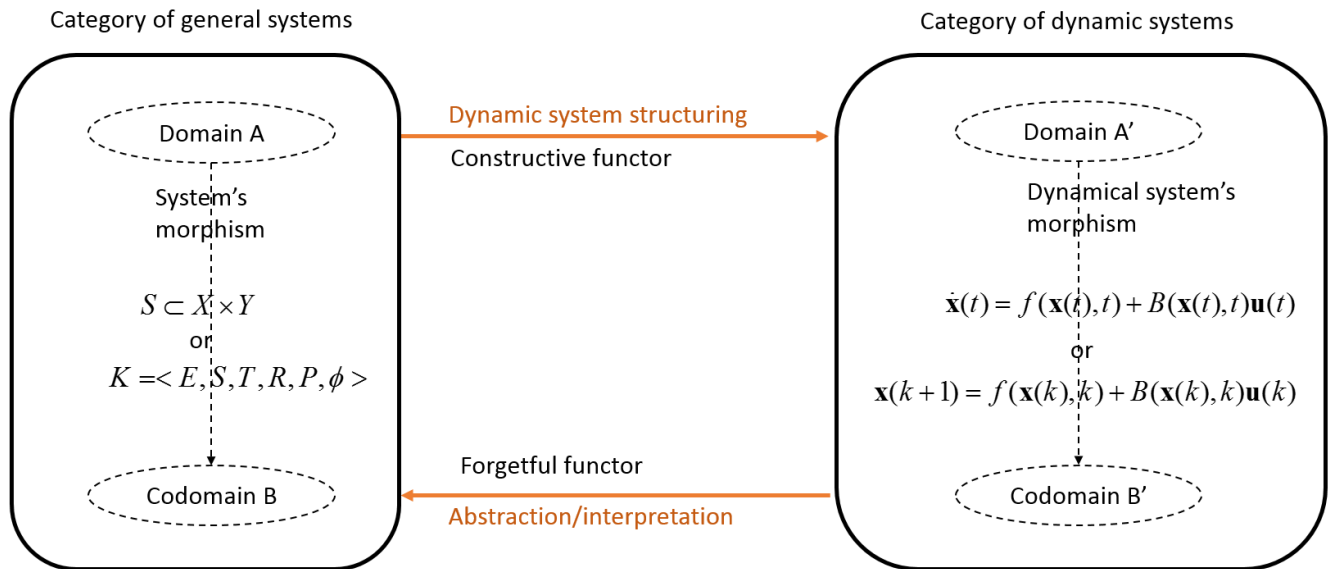
A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE Access

FIGURE 2. The role of category theory in mapping dynamic systems representations of general systems.

is the agent that makes its inquires on the *world* through its actuators on thermal devices. A temperature sensor gives timely the answer to this cyber-physical query to the world. To such a stable thermostat system, let us now associate a dynamic system that expresses its purposeful and stable behaviour; this operation is expressed in Figure 3a. The thermostat's "hands" are the on-off actions on the heater systems; its "eyes" are the temperature sensor; its "mind" are the regulation rules it applies. Note also that the functors between symbolic and dynamic representations can include, as particular cases, processes like defuzzification and fuzzification.

What happens if "unknown" arrives as input? This unknown token might be the representation of many issues in the physical world that the thermostat does not know how to handle. But, this also implies that the system has been designed with unexpected inside, leaving open and unspecified parts for the "unknown" case [94]. It might be a problem in the protocol of sensor acquisition, a fault, or breakage of some components. The original program is not able to describe this unexpected situation. It needs asking help to a meta-system that knows how to, identify, solve, and restore a new situation into a stable condition. The act of *knowing* is the query that the thermostat poses to the world. When the query does not provide consistent answer – for example something unexpected or unexplainable happens to the agent's worldview (its autospace) – then it is time for the agent to react in order to survive the situation and reorganize itself, searching for something else, through autonomous learning.

The capability of asking help, and so of switching to a higher representation of reality, is provided by endowing the agent with holonic behavior. In Figure 3b, the new stable condition is obtained by a change in the program that lets the thermostat wait for maintenance. The change in the program
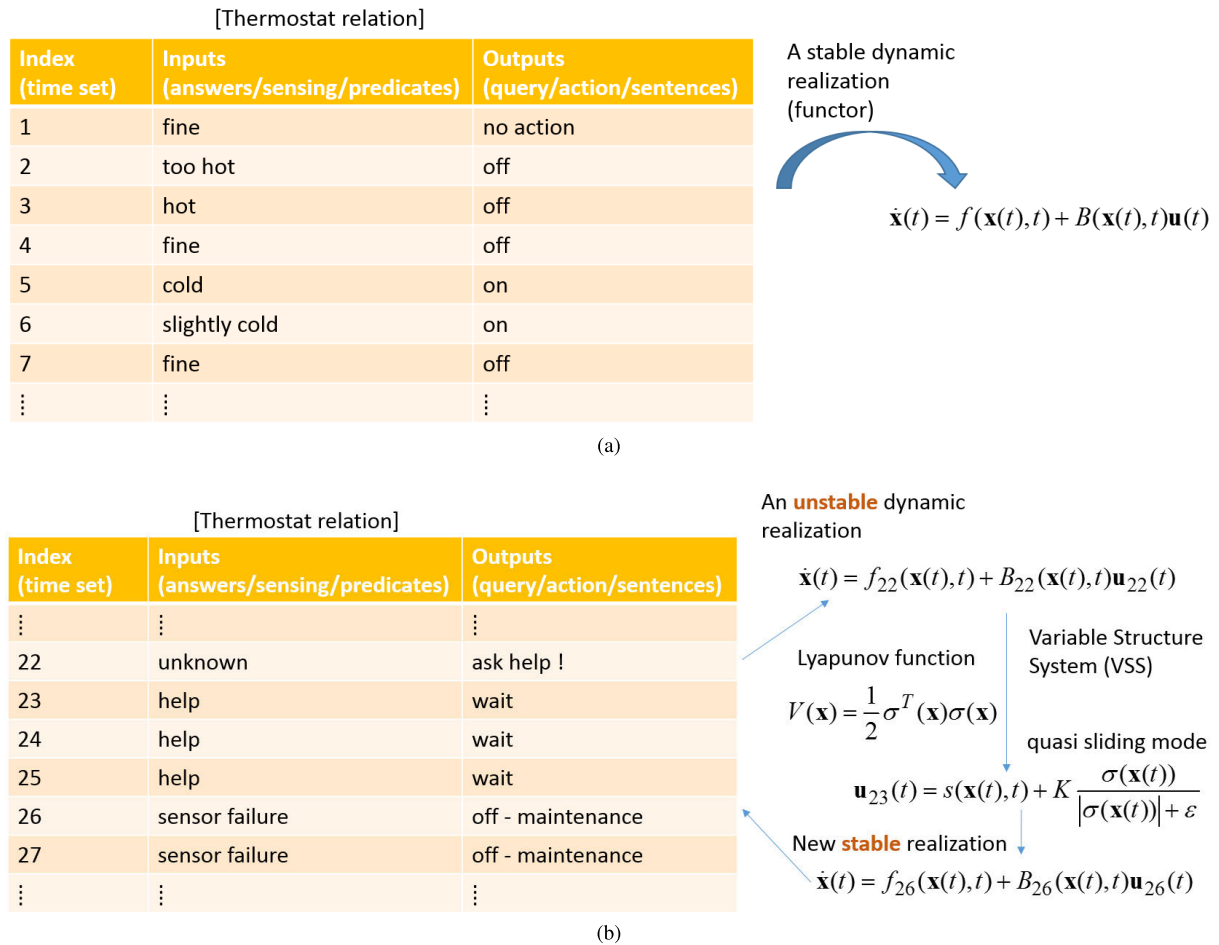
is obtained automatically if there is a control mechanism that brings the unstable dynamical system at time 22 ($\dot{\mathbf{x}}(t) = f_{22}(\mathbf{x}(t), t) + B_{22}(\mathbf{x}(t), t)\mathbf{u}_{22}(t)$) to start a process of stabilization. This process can start an automatic procedure that uses a quasi-sliding mode control (which notably is a rather general control technique) that forces the system to explore the state space (enlarge its worldview) towards a new stable condition, as described by a new dynamical system at time 26 ($\dot{\mathbf{x}}(t) = f_{26}(\mathbf{x}(t), t) + B_{26}(\mathbf{x}(t), t)\mathbf{u}_{26}(t)$). The new stable dynamical representation is then brought back into initial semantic realm, by the forgetful functor. The holonic agent implementing the forgetful functor issues a new program that keeps the thermostat in a maintenance request state; with a new functionality initially neither foreseen, neither specified, nor built-in.

The dynamical system transformation of this example shows paradigmatically that in principle any kind of symbolic or general system can be treated with well-established engineering tools that bring about some coherent and unified methodology for the realization of AI as solutions in complex reflexive scenario for agents.

The iterated application of robust control techniques, like sliding-mode techniques, allows navigating the multidimensional state space towards stable realities (sliding manifolds) as soon as an "autonomous" agent experiences "irritation" for an external unexpected perturbation of reality, as shown in Figure 4a.

When the system is endowed with a cybernetic property (by definition, a circular-causal and feedback mechanism in systems), then the relationship between *knowing* and *being* can constitute a transformation: the eigenfunction. The eigenvalues of the eigenfunction determine the centre of the dynamic system's region of stability.

IEEE *Access*

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

[Thermostat relation]

Index (time set)	Inputs (answers/sensing/predicates)	Outputs (query/action/sentences)
1	fine	no action
2	too hot	off
3	hot	off
4	fine	off
5	cold	on
6	slightly cold	on
7	fine	off
⋮	⋮	⋮

A stable dynamic realization (functor)

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t) + B(\mathbf{x}(t), t)\mathbf{u}(t)$$

(a)

[Thermostat relation]

Index (time set)	Inputs (answers/sensing/predicates)	Outputs (query/action/sentences)
⋮	⋮	⋮
22	unknown	ask help !
23	help	wait
24	help	wait
25	help	wait
26	sensor failure	off - maintenance
27	sensor failure	off - maintenance
⋮	⋮	⋮

An **unstable** dynamic realization

$$\dot{\mathbf{x}}(t) = f_{22}(\mathbf{x}(t), t) + B_{22}(\mathbf{x}(t), t)\mathbf{u}_{22}(t)$$

Variable Structure System (VSS)

Lyapunov function

$$V(\mathbf{x}) = \frac{1}{2}\sigma^T(\mathbf{x})\sigma(\mathbf{x})$$

quasi sliding mode

$$\mathbf{u}_{23}(t) = s(\mathbf{x}(t), t) + K\frac{\sigma(\mathbf{x}(t))}{|\sigma(\mathbf{x}(t))| + \varepsilon}$$

New **stable** realization

$$\dot{\mathbf{x}}(t) = f_{26}(\mathbf{x}(t), t) + B_{26}(\mathbf{x}(t), t)\mathbf{u}_{26}(t)$$

(b)

FIGURE 3. (a) An intelligent thermostat is an agent which stable regulating behaviour can be expressed through a dynamical system. (b) A new control law (quasi-sliding mode) obtains a new stable program.

Note also that, in this example, the notion of optimal control or optimal operational search has not been specified. This has been a deliberate decision as the scenario of interest here is complexity. In the context of complexity, the first deal for the intelligent agent is to maintain organization and a state of viability. After this accomplishment, a continuous improvement teleology (in cybernetics sense) of the holarchy can be desirable, though not strictly necessary. Monotonic improvement of performance can be superimposed on the holarchy in form of a management goal as treated, for example, in the holonic management tree methodology [73], [111]. Optimality is a next step, but it cannot be thought of as always available and viable in complex scenarios.

In Figure 4b, the path to the solution of a problem can be draft with a suitable covering of formal partially overlapped systems (Sk, with $k = 1, \ldots, 8$). If we attributed some topological property to Sk (e.g., Hausdorff separability), the more is the distance from the centre, where the system is stable, the more the agent starts to feel uncomfortable, and to be less aware on how to control things. Indeed, if we use a Lyapunov criteria to determine the stability point of a closed loop control of a *being-and-knowing* system for the agent, the

locus of stability will be assigned naturally as the center of the system.

During problem solution, it might be necessary to switch from one reality to another. When the edge of a formal system is reached, the formal system cannot avoid paradoxes (inconsistency) and so undecidability in computational terms. It happens in particular in ill-defined and complex problems where no formal system can guarantee alone a solution, as treated in section III-A. The overlaying of holarchy and sliding-mode control is a possible step towards the harnessing of problems for autonomous agents in complex environment.

3) RMAS, A MULTI-AGENT TECHNOLOGY FOR A REALIZATION OF THE APPROACH

In order to get a suitable implementation that satisfies the requirements of the bottom-up approach to AGI, the RMAS (relational-model multi-agent system) architecture in [13] and [15] is proposed. RMAS is able to integrate algorithmic computation, holonic eigenforms, lambda calculus, and other computational means, in particular any object that monads in category theory can handle. A solid formalism of the model

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE*Access*
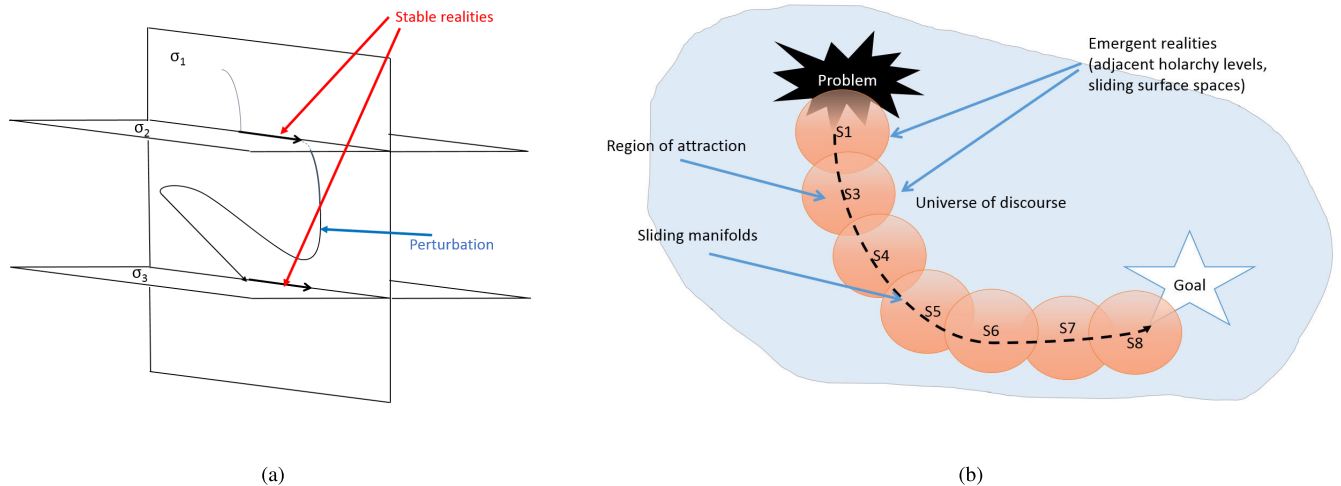
FIGURE 4. (a) The switch of context into stable realities for the agent through sliding manifolds. (b) Purposeful crossing of formal systems by self-organizing holarchy towards a goal.

of computation of RMAS has been recently established in order to harness and fully express its possibilities, in particular in bridging the disciplines of control and systems engineering and computer science, whilst concerning artificial intelligence [14]. With these features, RMAS represents a practicable link and convergence viable means between the bottom-up and the top-down approaches towards AGI, as discussed in this work.

The core of the RMAS architecture rests on the use of the *Relational Model* (RM). The RM (*Relational Model*) intended in RMAS is an evolution of Codd's relational model [112], which later developments have been obtained by Date and Darwen [113]. The essence of RM, and its most favorable feature to our purposes, is that all the queries and manipulative operators rely upon *relations*, and all of them generate *relations* as a result [112]. In essence, only one recursive universal type is needed in RM. This is a key feature when we have to handle holarchies spanning indefinite objects: RM is invariant with respect to the semantic content of its relations. RM is relational algebra with an associated choice of relational language (RL), which fundamental set of features are the following.

- Any operand in the relational algebra is a relation variable (*relvar*).
- A *relvar* is any relational object with no specific type a-priori defined, e.g. a whole database, a table, a catalog, a schema etc. In standard database management systems those objects are managed by vendor-specific language extensions (e.g. command to query the list of tables in a database).
- The type of any operand in the relational algebra is defined recursively, and possibly only when needed, in a typical lazy or low-and-late approach – designed for the unexpected.

The RMAS is based on the integration of two main guiding paradigms: the Relational Model (RM) and oracular compu-

tation (OC) [114]. Oracular computation is expressed pictorially in Figure 5. An oracle is a black-box object (with a well-defined set of input-output relations) that is queried from the algorithmic part of the program (formerly referred as the act of *knowing*). These black-boxes can be any other model of computation, being it any Turing machine, a huge distributed database, a neural network, an analog circuit, or any other of the computations today relevant for AI.

Nowadays, AI is using a plethora of computation models ranging from neat formal digital methods to scruffy methods, like Bayesian or deep neural networks [115]. Other promising methods can rely for example on morphological computation and reservoir computing [116], morphic computing [117], and quantum computing for machine learning [114]. Other researchers developed recently a prototype of another kind of computation that uses inverse problem solving method, genetic algorithm, and cognitive and quantum semantics as tools for collective intelligence, involving networked expertise support, convergent decision making with cognitive modelling, and cognitive models of verification by mapping relevant Big Data [28].

RMAS is an experimental framework for RM that aims first at achieving a suitable coupling between computational languages and cyber-physical resources, in order to explore the lands of autonomic computing for multi-agent systems, towards real autonomy in artificial machines [13]. RMAS wants to constitute viable experimental means for software capable of improving itself, which has been the dream of computer scientists since the inception of the AI field. By following Yampolskiy [42], Recursive Self-Improvement (RSI) is the only type of improvement that has the potential to completely replace the original algorithm with a completely different one. This mechanism can be enabled by the dynamical type construction and the inherently endofunctorial structure of the RM.

The RMAS has been used to express the capabilities of autonomic systems, in compliance with industrial standards
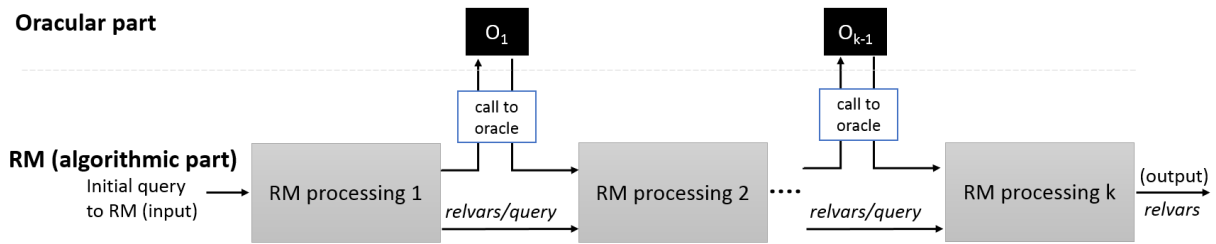
Oracular part

FIGURE 5. Schematic architecture of the connection between oracular computation and the algorithmic part (RM) in the RMAS architecture.

(like IEC 61499) [118], with the favorable added feature of scalability from the Cloud computing down to very tiny (swarms of) embedded systems [96].

IV. REAL APPLICATIONS AND DISCUSSION

The real practice of using components of the convergent AI resonator has a 30-years long history, but with a separated action between bottom-up and top-down directions. Only recently a convergence has started [83], [119], [120]. At the basis of cognition in humans and machines, it is expected that the collision or resonance of the two approaches, coming from the bottom and the top of the intelligence phenomenon, will reveal new unexplored but decisive features.

For the bottom-up part, experiments on the RMAS architecture are currently facing the complexity of self-programming and autonomic capabilities of computing machines [13]. The main experimental limit currently existing is represented by the unavailability of a suitable implementation of the Relational Model. This difficulty is nevertheless partially attenuated by the RMAS architecture itself. It allows the coupling of relational database management systems (as best available technologies) with any other form of computation.

Looking at Figure 5, the algorithmic part can be considered as constituted, in practical terms, from a combination of computing languages. An example of effective combination is the SQLite (database manipulation language) plus *Haskell*, a language that can conveniently handle monads, and functors of category theory. *Haskell* has a functional and *lazy* execution model. It means that the type of one object is actually defined only when strictly needed. This feature allows operations across several levels of abstractions whilst keeping invariant the relational structure of the calculus at many different levels of reality (i.e., domains of semantic interpretation); it is the relation between the objects that dominates with respect to their details – the form not the substance. The focus is on patterns rather than data and their domain-dependent types. In addition, the oracular part in RMAS admits the use of some non-algorithmic means like neuromorphic computing, deep learning, analog computing, quantum computing, and holography. This permits homogeneous relational structures to emerge from heterogeneous computations and so objects coming from completely different semantic realities.

In practical applications, RMAS yearns for becoming a suitable probe, in order to monitor and study in more depth the encounter between cognitivist and connectivist approaches, in a mostly explainable ground.

On the other side, for the top-down part, a most important experimentation should explore hybrid organizations of human and machines that can interact at the relational level. With appropriate interfaces that basically exchange information through queries, it is possible to pin down the information and the operations in an understandable form for the actors involved, and at the suitable level of knowledge required from the discourse.

An example of top-down process that ensures the convergence of humans and multi-agent interaction is the application in real practice of the authors' method of manufacturer's strategic risk temperature assessment with cognitive modelling and Blockchain technology [121]. Strategic planning in manufacturing management is a multi-level and ill-defined decision-making process that is characterized by high strategic risks. Many implements, participants, and factors influence the definition goals and paths towards their achievement in manufacturing. All data from implements have to be taken into account; the participants must take into account their non-formalised interests and desires; the numbers of factors may arrive to more than one hundred. The convergent approach, including cognitive modelling, genetic algorithm and trust space, created with Blockchain, can accelerate assessing and reduce strategic risks. The three blocks each of which contains 8 parameters have to be evaluated.

The approach has been tested in real practice in the field of city building. It was also applied in the case of the strategic risk temperature assessment for the plants which admission is strictly limited. The strategic risk assessment can be enforced by top management of the enterprise; when tested it took 1-2 hours. The first proposals of steps for improving the strategic manufacturers' management took less than 1 week.

Convergent approach with a special structuring of information during meetings allowed to significantly accelerating their implementation [119], [121]. Cognitive semantics was taken into account in an indirect way. To enhance the inclusion of cognitive semantics in a weakly structured AI modelling, the convergent approach have been used with cognitive modelling [119] and some concepts of category theory [83].

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE Access

For example, the creation of the Megapolis (Moscow, Russia) tourism strategic planning required a baseline of 35 brainstorming sessions [6]. The strategic analysis was made and a cognitive model was built. The new process took 4 hours. The strategic planning is a poorly formalized process with a typical top-down approach. It is realized by top management of the company or government's department. The process is unique and is represented by many conceptual characteristics (factors). The information is unreliable. In the strategic decision-making process it is necessary to take into account such not easily formalizable factors as institutional constructions, human resource, socio-cultural environment, emotions, feelings, latent interests of the participants. During convergent brainstorming more than 65 factors were generated, then the substantial 15 factors were selected and the importance of the factors and their interaction were evaluated. The cognitive model was created with its verification by mapping in the relevant sets of Big Data.

The next example of top-down applications is in the field of accelerating the processes of networked democracy [28] or increasing corporate responsibility. These processes embrace the participants' activity of different levels of control — from folk or employers to top-management or government. The top-down and bottom-up activity have to be reconciled. The business and social acting have to be convergent with taking into account heterogeneous factors in an ethical and transparent way. Author's convergent methodology ensures the integrity, purposefulness and sustainability of developments of collectives in the external environment.

The most difficult part in creating AI systems is taking into account the cognitive dimension of the interaction of the various components of control systems. Such components are: the upper and lower levels of control, formalizable and non-formalizable semantics of AI models, the uncaused decision making, the interface of a machine and a person, the ethical and unconscious aspects of decision-making, and so on. These components "speak" different languages, and have different cognitive semantic interpretations. Often, from the description of the problem of a certain subject area, the need and the possibility of using AI to support decision-making is not obvious. This is largely due to the complete non-formalizability and causeless of the cognitive components of the collective decision-making process: thoughts, feelings, and cognitive processes of participants.

It can be hypothesized that such problems can be solved using Strong AI and AGI approaches. At the same time, actual AI achievements tell us that only bespoke solutions are available, still far from AGI or Strong AI targets.

The proposed encounter of top-down and bottom-up approaches could be the next steps in the developing of cognitive aspects of AI that will be transforming it in a new hybrid AI technology.

In the context of Hybrid AI development, ethics is a rather dangerous area for discussion: it is being broken and curtailed. The following aspects and questions characterize ethics:

- the results of the long evolution of humankind, ancient heritage and postmodernism,
- types of ontologies,
- posing such questions as "The value of human life? What is justice?"
- ethical diversity.

As a result of the impact of the digital environment and AI, the following occurs:

- hacking of the psyche and consciousness,
- the risks of hybridization of cybernetic and natural spheres are growing,
- the evolution of morality leads to the fall of its universal principles.

V. CONCLUSION

In this work, a path towards new insights into the implications that are provoked by the existing duality in cognitive aspects of humans and machines has begun. A first implication of such a duality is in the need of control of the unavoidable collision at the contact point between two different worlds: that of the non-formalizable objects and behaviours, and that of formalizable tokens of reality that still can be successfully, though partially, controlled with traditional AI approaches.

In order to provide such a path, the research objective of this paper was to examine the interesting zone that can be generated when completely different approaches to AI are confronted and integrated.

The first technological approach here considered, is that of a non-formalizable cognitive and organizational AI methodology that was proven effective in the acceleration of decision-making processes. This approach usually proceeds top-down from non-formalizable human consciousness and unconsciousness elements of reality towards practicable tools of AI. The second technological approach considered proceeds in opposite direction, from the bottom of current state-of-the-art artificial cognitive and AI technology up to a human-level capability of autonomous representation and reasoning.

The main result that this work has targeted is the creation of a research and development framework that can handle and leverage the pros and cons of both the aforementioned approaches in order to provide a viable direction towards next levels of cognitive human-machine systems.

In order to achieve that ambitious objective, this paper started with a thorough overview of cognitive architectures and of current trends in AI developments, which could constitute a solid reference for the new paper propositions. In particular, a position was made on the definition of an appropriate set of interdisciplinary tools (theoretical, methodological, and technological) that, under a new interpretation of constructivism, allows creating a fertile ground for the development of practical examples of general artificial intelligence (AGI). The examples here provided aimed at the introduction and at the coupling between the convergent top-down approach for decision-making processes and some seeds of autonomic

computational agents, which are expected to progress AI research further, possibly towards concepts and visions of Strong AI in their future developments.

The novelty of this research work comes from the leveraging of two complementary and state-of-the-art research paths from the authors that are fronted and tentatively synergized for the first time. It aims to making the behaviour of the manifold human-machine systems much more convergent, controlled, purposeful and sustainable. The developments implied by this research should constitute a new contribution to the field of collective human-machine decision-making with AI in different segments of the digital economy and stimulate transdisciplinary debate and knowledge for socio-humanitarian and managerial aspects.

This work determines the extent of dualities of different models in AI that confront cognitivist and connectionist architectures, conscious and unconscious action, symbolic and conceptual realizations, emergent and brain-based computing, automata and humans, quantity and quality factors, formalized and non-formalized semantics. Some of these frameworks may come in form of algorithms or requirements, rules of strategic conversations or network brainstorming, and quantum or relativistic semantics.

With the moral diversity of different fragments of society, equally diverse ethics are embedded in AI systems. However, due to the digital degradation of thinking and the digital distortion of ethics, the aspects of social responsibility are still poorly perceived and powerless against the business incentives of the market.

The discussion here proposed can be used to provide perspectives on the shaping of the societal, ethical, and normative impacts of the symbiosis of humans and machines. It has been shown that there is a huge research concerning the study of human-machine duality and symbiosis, together with models that try to norm the frameworks in AI. These studies have deep implications in the public administration, production development, network democracy processes, elaboration of societal rules and of human ethics, and are prone to define with ultimate clarity the norms for purposeful and sustainable use of humans and machines.

The limits of the proposed methodological position are inherent in the needs of a consistent research effort in order to substantiate the claims and the visions here provided, and to express an urgent involvement of the AI research community on these themes, which would give strength to our proposed work plan. Future work from authors will be oriented to the design and the conduction of appropriate experimental experience on the aspects involved by this positional work.

REFERENCES

[1] A. Raikov, "Post-non-classical artificial intelligence and its pioneer practical applications," *IFAC-PapersOnLine*, vol. 52, no. 25, pp. 343–348, 2019, doi: 10.1016/j.ifacol.2019.12.547.

[2] S. Tolan, A. Pesole, F. Martínez-Plumed, E. Fernández-Macías, J. Hernández-Orallo, and E. Gómez, "Measuring the occupational impact of AI: Tasks, cognitive abilities and AI benchmarks," *J. Artif. Intell. Res.*, vol. 71, pp. 191–236, Jun. 2021, doi: 10.1613/jair.1.12647.

[3] M. Alfonseca, M. Cebrian, A. Fernandez Anta, L. Coviello, A. Abeliuk, and I. Rahwan, "Superintelligence cannot be contained: Lessons from computability theory," *J. Artif. Intell. Res.*, vol. 70, pp. 65–76, Jan. 2021, doi: 10.1613/jair.1.12202.

[4] L. Cheng, K. R. Varshney, and H. Liu, "Socially responsible AI algorithms: Issues, purposes, and challenges," *J. Artif. Intell. Res.*, vol. 71, pp. 1137–1181, Aug. 2021, doi: 10.1613/jair.1.12814.

[5] D. Gubanov, N. Korgin, D. Novikov, and A. Raikov, *E-Expertise: Modern Collective Intelligence*. Cham, Switzerland: Springer, 2014.

[6] A. Raikov, "Accelerating decision-making in transport emergency with artificial intelligence," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 6, pp. 520–530, Nov. 2020, doi: 10.25046/aj050662.

[7] W. J. Clancey, "Situated action: A neuropsychological interpretation response to Vera and Simon," *Cogn. Sci.*, vol. 17, no. 1, pp. 87–116, 1993, doi: 10.1016/S0364-0213(05)80013-8.

[8] W. J. Clancey, *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[9] V. Raja, "A theory of resonance: Towards an ecological cognitive architecture," *Minds Mach.*, vol. 28, no. 1, pp. 29–51, Mar. 2018, doi: 10.1007/s11023-017-9431-8.

[10] S. T. Piantadosi, "The computational origin of representation," *Minds Mach.*, vol. 31, no. 1, pp. 1–58, Mar. 2021, doi: 10.1007/s11023-020-09540-9.

[11] P. Wang, X. Li, and P. Hammer, "Self in NARS, an AGI system," *Frontiers Robot. AI*, vol. 5, p. 20, Mar. 2018, doi: 10.3389/frobt.2018.00020.

[12] K. R. Thórisson, "A new constructivist AI: From manual methods to self-constructive systems," in *Theoretical Foundations of Artificial General Intelligence*. Cham, Switzerland: Springer, 2012, pp. 145–171, doi: 10.2991/978-94-91216-62-6_9.

[13] A. Bonci, S. Longhi, and M. Pirani, "RMAS architecture for autonomic computing in cyber-physical systems," in *Proc. IECON 45th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2019, pp. 2996–3003, doi: 10.1109/IECON.2019.8926991.

[14] M. Pirani, A. F. Dragoni, and S. Longhi, "Towards sustainable models of computation for artificial intelligence in cyber-physical systems," in *Proc. IECON 47th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2021, pp. 1–8, doi: 10.1109/IECON48115.2021.9589594.

[15] M. Pirani, A. Bonci, and S. Longhi, "Towards a formal model of computation for RMAS," *Proc. Comput. Sci.*, vol. 200, pp. 865–877, Jan. 2022, doi: 10.1016/j.procs.2022.01.284.

[16] C. Lewry, K. Curtis, N. Vasilyeva, F. Xu, and T. L. Griffiths, "Intuitions about magic track the development of intuitive physics," *Cognition*, vol. 214, Sep. 2021, Art. no. 104762, doi: 10.1016/j.cognition.2021.104762.

[17] H. Lyre, "The state space of artificial intelligence," *Minds Mach.*, vol. 30, no. 3, pp. 325–347, Sep. 2020, doi: 10.1007/s11023-020-09538-3.

[18] S. Sundvall, "Artificial intelligence," in *Critical Terms in Futures Studies*, P. H. Winston, Ed. Cham, Switzerland: Palgrave, 2019, pp. 29–34.

[19] I. Perko, "Hybrid reality development—Can social responsibility concepts provide guidance?" *Kybernetes*, vol. 50, no. 3, pp. 676–693, Mar. 2021, doi: 10.1108/K-01-2020-0061.

[20] R. Fjelland, "Why general artificial intelligence will not be realized," *Humanities Social Sci. Commun.*, vol. 7, no. 1, pp. 1–9, Dec. 2020, doi: 10.1057/s41599-020-0494-4.

[21] *Artificial Intelligence Concepts and Terminology*, document ISO/IEC CD 22989, Working Draft, 2019.

[22] E. Esposito, "Artificial communication? The production of contingency by algorithms," *Zeitschrift Für Soziologie*, vol. 46, no. 4, pp. 249–265, Aug. 2017.

[23] J. Peregrin, "Do computers 'have syntax, but, no semantics'," *Minds Mach.*, vol. 31, no. 2, pp. 305–321, Jun. 2021, doi: 10.1007/s11023-021-09564-9.

[24] A. Azizi, "Hybrid artificial intelligence optimization technique," in *Applications of Artificial Intelligence Techniques in Industry 4.0* (SpringerBriefs in Applied Sciences and Technology). Singapore: Springer, 2019, pp. 27–47, doi: 10.1007/978-981-13-2640-0_4.

[25] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, 2020, doi: 10.1145/3386252.

[26] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds Mach.*, vol. 30, pp. 681–694, Nov. 2020, doi: 10.1007/s11023-020-09548-1.

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE *Access*

[27] D. Widdows, K. Kitto, and T. Cohen, "Quantum mathematics in artificial intelligence," *J. Artif. Intell. Res.*, vol. 72, pp. 1307–1341, Dec. 2021, doi: 10.1613/jair.1.12702.

[28] A. Raikov, "Accelerating technology for self-organising networked democracy," *Futures*, vol. 103, pp. 17–26, Oct. 2018, doi: 10.1016/j. futures.2018.03.015.

[29] A. C. Neubauer, "The future of intelligence research in the coming age of artificial intelligence—With a special consideration of the philosophical movements of trans- and posthumanism," *Intelligence*, vol. 87, Jul. 2021, Art. no. 101563, doi: 10.1016/j.intell.2021.101563.

[30] U. Straccia, *Foundations of Fuzzy Logic and Semantic Web Languages*. Oxfordshire, U.K.: Taylor & Francis, 2013.

[31] H. van Ditmarsch, T. French, F. R. Velázquez-Quesada, and Y. N. Wáng, "Implicit, explicit and speculative knowledge," *Artif. Intell.*, vol. 256, pp. 35–67, Mar. 2018, doi: 10.1016/j.artint.2017.11.004.

[32] M. Gavanelli, E. Lamma, F. Riguzzi, E. Bellodi, R. Zese, and G. Cota, "Reasoning on datalog± ontologies with abductive logic programming," *Fundam. Inform.*, vol. 159, nos. 1–2, pp. 65–93, 2018, doi: 10.3233/FI-2018-1658.

[33] C. P. Boje, "Knowledge representation, storage and retrieval for BIM supported building evacuation design," Ph.D. dissertation, BRE Trust Centre Sustain. Eng., Cardiff Univ., Cardiff, Wales, 2018. [Online]. Available: https://orca.cardiff.ac.uk/id/eprint/112966/

[34] B. Naticchia, L. Messi, M. Pirani, A. Bonci, A. Carbonari, and L. C. Tolve, "Holonic system for real-time emergency management in buildings," in *Proc. Int. Symp. Autom. Robot. Construction (IAARC)*, May 2019, pp. 453–460, doi: 10.22260/ISARC2019/0061.

[35] P. Masters and S. Sardina, "Expecting the unexpected: Goal recognition for rational and irrational agents," *Artif. Intell.*, vol. 297, Aug. 2021, Art. no. 103490, doi: 10.1016/j.artint.2021.103490.

[36] S. Hafenbrädl, D. Waeger, J. N. Marewski, and G. Gigerenzer, "Applied decision making with fast-and-frugal heuristics," *J. Appl. Res. Memory Cognition*, vol. 5, no. 2, pp. 215–231, Jun. 2016, doi: 10.1016/j.jarmac.2016.04.011.

[37] J. S. Mill, *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*, 8th ed. New York, NY, USA: Harper & Brothers, 1882.

[38] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," 2018, *arXiv:1801.04016*.

[39] P. K. Mistry, E. M. Pothos, J. Vandekerckhove, and J. S. Trueblood, "A quantum probability account of individual differences in causal reasoning," *J. Math. Psychol.*, vol. 87, pp. 76–97, Dec. 2018, doi: 10.1016/j.jmp.2018.09.003.

[40] M. Setzu, R. Guidotti, A. Monreale, F. Turini, D. Pedreschi, and F. Giannotti, "GLocalX—From local to global explanations of black box AI models," *Artif. Intell.*, vol. 294, May 2021, Art. no. 103457, doi: 10.1016/j.artint.2021.103457.

[41] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behav. Brain Sci.*, vol. 40, pp. 1–72, Apr. 2017, doi: 10.1017/S0140525X16001837.

[42] R. V. Yampolskiy, "From seed AI to technological singularity via recursively self-improving software," 2015, *arXiv:1502.06512*.

[43] K. R. Thórisson, "Seed-programmed autonomous general learning," in *Proc. Int. Workshop Self-Supervised Learn.*, 2020, pp. 32–61.

[44] I. Kotseruba and J. K. Tsotsos, "40 years of cognitive architectures: Core cognitive abilities and practical applications," *Artif. Intell. Rev.*, vol. 53, no. 1, pp. 17–94, 2018, doi: 10.1007/s10462-018-9646-y.

[45] P. Wang, "Natural language processing by reasoning and learning," in *Proc. Int. Conf. Artif. General Intelligence*, Cham, Switzerland: Springer, 2013, pp. 160–169, doi: 10.1007/978-3-642-39521-5_17.

[46] U. Faghihi and S. Franklin, "The LIDA model as a foundational architecture for AGI," in *Theoretical Foundations of Artificial General Intelligence. Atlantis Thinking Machines*, vol. 4, P. Wang and B. Goertzel, Eds. Paris, France: Atlantis Press, 2012, pp. 103–121, doi: 10.2991/978-94-91216-62-6_7.

[47] S. Schaat, A. Wendt, S. Kollmann, F. Gelbard, and M. Jakubec, "Interdisciplinary development and evaluation of cognitive architectures exemplified with the SiMA approach," in *Proc. EAPCogSci*, Sep. 2015, pp. 25–27.

[48] D. V. Pynadath, P. S. Rosenbloom, and S. C. Marsella, "Reinforcement learning for adaptive theory of mind in the sigma cognitive architecture," in *Proc. Int. Conf. Artif. General Intell.*, Cham, Switzerland: Springer, 2014, pp. 143–154, doi: 10.1007/978-3-319-09274-4_14.

[49] W. Bridewell and P. Bello, "Incremental object perception in an attention-driven cognitive architecture," in *Proc. 37th Annu. Meeting Cognit. Sci. Soc.*, 2015, pp. 279–284.

[50] U. Faghihi, P. Poirier, and O. Larue, "Emotional cognitive architectures," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Berlin, Germany: Springer, 2011, pp. 487–496, doi: 10.1007/978-3-642-24600-5_52.

[51] N. Chomsky, "Language and nature," *Mind, New Ser.*, vol. 104, no. 413, pp. 1–61, 1995.

[52] A. Tozzi and D. Papo, "Projective mechanisms subtending real world phenomena wipe away cause effect relationships," *Prog. Biophys. Mol. Biol.*, vol. 151, pp. 1–13, Mar. 2020, doi: 10.1016/j.pbiomolbio. 2019.12.002.

[53] M. Eisenberg and R. Guy, "A proof of the hairy ball theorem," *Amer. Math. Monthly*, vol. 86, no. 7, pp. 571–574, Aug. 1979.

[54] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" *Phys. Rev.*, vol. 47, pp. 777–780, May 1935.

[55] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[56] The BIG Bell Test Collaboration, "Challenging local realism with human choices," *Nature*, vol. 557, no. 7704, p. 212, 2018. [Online]. Available: https://www.nature.com/articles/s41586-018-0085-3, doi: 10.1038/s41586-018-0085-3.

[57] H. L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence*. Manhattan, NY, USA: Harper & Row, Harper Colophon Books, 1979.

[58] T.-J. Huang, "Imitating the brain with neurocomputer a 'new' way towards artificial general intelligence," *Int. J. Autom. Comput.*, vol. 14, no. 5, pp. 520–531, Oct. 2017, doi: 10.1007/s11633-017-1082-y.

[59] C. Villani, L. Lugli, M. T. Liuzza, R. Nicoletti, and A. M. Borghi, "Sensorimotor and interoceptive dimensions in concrete and abstract concepts," *J. Memory Lang.*, vol. 116, Feb. 2021, Art. no. 104173, doi: 10.1016/j.jml.2020.104173.

[60] H. R. Maturana and F. J. Varela, *The Tree of Knowledge: The Biological Roots of Human Understanding*. New Delhi, India: New Science Library/Shambhala Publications, 1987.

[61] A. Newen, L. D. Bruin, and S. Gallagher, *The Oxford handbook of 4E Cognition*. London, U.K.: Oxford Univ. Press, 2018.

[62] R. G. Licata I, "Computation by intention and electronic image of the brain," *J. Appl. Comput. Math.*, vol. 4, no. 4, pp. 232–244, 2015, doi: 10.4172/2168-9679.1000232.

[63] W. J. Freeman, "Nonlinear neurodynamics of intentionality," *J. Mind Behav.*, vol. 18, nos. 2–3, pp. 291–304, 1997.

[64] D. Chalmers, "The hard problem of consciousness," in *The Blackwell Companion to Consciousness*, S. Schneider and M. Velmans, Eds. Hoboken, NJ, USA: Wiley, 2017, pp. 32–42, doi: 10.1002/9780470751466.ch18.

[65] H. Atmanspacher, "Quantum approaches to brain and mind. An overview with representative examples," in *The Blackwell Companion to Consciousness*, S. Schneider and M. Velmans, Eds. Hoboken, NJ, USA: Wiley, 2017, pp. 298–313, doi: 10.1002/9781119132363.ch21.

[66] A. Dalela, *Quantum Meaning: A Semantic Interpretation Quantum Theory*. Bangalore, India: Shabda Press, 2012.

[67] G. Gigerenzer, *Gut Feelings. The Intelligence of the Unconscious*. London, U.K.: Viking, 2007.

[68] D. N. Perkins, *The Eureka Effect: The Art and Logic of Breakthrough Thinking*. New York, NY, USA: WW Norton & Company, 2001.

[69] A. Bonci, M. Pirani, A. Cucchiarelli, A. Carbonari, B. Naticchia, and S. Longhi, "A review of recursive holarchies for viable systems in CPSs," in *Proc. IEEE 16th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2018, pp. 37–42, doi: 10.1109/INDIN.2018.8472055.

[70] A. Raikov, "Convergent networked decision-making using group insights," *Complex Intell. Syst.*, vol. 1, nos. 1–4, pp. 57–68, Dec. 2015, doi: 10.1007/s40747-016-0005-9.

[71] R. Dobbe, T. K. Gilbert, and Y. Mintz, "Hard choices in artificial intelligence," *Artif. Intell.*, vol. 300, Nov. 2021, Art. no. 103555, doi: 10.1007/s10845-020-01532-x.

[72] R. Dazeley, P. Vamplew, C. Foale, C. Young, S. Aryal, and F. Cruz, "Levels of explainable artificial intelligence for human-aligned conversational explanations," *Artif. Intell.*, vol. 299, Oct. 2021, Art. no. 103525, doi: 10.1016/j.artint.2021.103525.

[73] A. Bonci, S. Longhi, and M. Pirani, "Holonic management tree technique for performance improvement over self-similar system structures," *Manage. Stud.*, vol. 7, no. 5, pp. 379–413, May 2019, doi: 10.17265/2328-2185/2019.05.001.

[74] M. Ienca, "Democratizing cognitive technology: A proactive approach," *Ethics Inf. Technol.*, vol. 21, no. 4, pp. 267–280, Dec. 2019, doi: 10.1007/s10676-018-9453-9.

[75] J. Maclure, "AI, explainability and public reason: The argument from the limitations of the human mind," *Minds Mach.*, vol. 31, no. 3, pp. 421–438, Sep. 2021, doi: 10.1007/s11023-021-09570-x.

[76] T. D. W. Wilcockson, E. M. Pothos, A. M. Osborne, and T. J. Crawford, "Top-down and bottom-up attentional biases for smoking-related stimuli: Comparing dependent and non-dependent smokers," *Addictive Behav.*, vol. 118, Jul. 2021, Art. no. 106886, doi: 10.1016/j.addbeh.2021.106886.

[77] M. Constantinescu, C. Voinea, R. Uszkai, and C. Vică, "Understanding responsibility in responsible AI. Dianoetic virtues and the hard problem of context," *Ethics Inf. Technol.*, vol. 23, no. 4, pp. 803–814, Dec. 2021, doi: 10.1007/s10676-021-09616-9.

[78] C. Allen, I. Smit, and W. Wallach, "Artificial morality: Top-down, bottom-up, and hybrid approaches," *Ethics Inf. Technol.*, vol. 7, no. 3, pp. 149–155, 2005, doi: 10.1007/s10676-006-0004-4.

[79] P. G. R. de Almeida, C. D. dos Santos, and J. S. Farias, "Artificial intelligence regulation: A framework for governance," *Ethics Inf. Technol.*, vol. 23, no. 3, pp. 505–525, Sep. 2021, doi: 10.1007/s10676-021-09593-z.

[80] C. Rovelli, "Relational quantum mechanics," *Int. J. Theor. Phys.*, vol. 35, no. 8, pp. 1637–1678, Aug. 1996, doi: 10.1007/BF02302261.

[81] A. Raikov, *Cognitive Semantics of Artificial Intelligence: A New Perspective* (Topics: Computational Intelligence XVII). Singapore: Springer, 2021, doi: 10.1007/978-981-33-6750-0.

[82] V. K. Ivanov, "Incorrect problems in topological spaces," *Siberian Math. J.*, vol. 10, no. 5, pp. 785–791, 1970, doi: 10.1007/BF00971654.

[83] A. N. Raikov, A. N. Ermakov, and A. A. Merkulov, "Assessments of the economic sectors needs in digital technologies," *Lobachevskii J. Math.*, vol. 40, no. 11, pp. 1837–1847, Nov. 2019, doi: 10.1134/S1995080219110246.

[84] H. A. Abbass, "Social integration of artificial intelligence: Functions, automation allocation logic and human-autonomy trust," *Cognit. Comput.*, vol. 11, no. 2, pp. 159–171, Apr. 2019, doi: 10.1007/s12559-018-9619-0.

[85] E. R. Palacios, A. Razi, T. Parr, M. Kirchhoff, and K. Friston, "On Markov blankets and hierarchical self-organisation," *J. Theor. Biol.*, vol. 486, Feb. 2020, Art. no. 110089, doi: 10.1016/j.jtbi.2019.110089.

[86] J. Winning and W. Bechtel, "Rethinking causality in biological and neural mechanisms: Constraints and control," *Minds Mach.*, vol. 28, no. 2, pp. 287–310, Jun. 2018, doi: 10.1007/s11023-018-9458-5.

[87] R. Orpwood, "Information and the origin of qualia," *Frontiers Syst. Neurosci.*, vol. 11, p. 22, Apr. 2017, doi: 10.3389/fnsys.2017.00022.

[88] C. Horsman, S. Stepney, R. C. Wagner, and V. Kendon, "When does a physical system compute?" *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 470, no. 2169, Sep. 2014, Art. no. 20140182, doi: 10.1098/rspa.2014.0182.

[89] S. C. Fletcher, "Computers in abstraction/representation theory," *Minds Mach.*, vol. 28, no. 3, pp. 445–463, Sep. 2018, doi: 10.1007/s11023-018-9470-9.

[90] J. Szangolies, "The abstraction/representation account of computation and subjective experience," *Minds Mach.*, vol. 30, no. 2, pp. 259–299, Jun. 2020, doi: 10.1007/s11023-020-09522-x.

[91] A. Koestler, *The Ghost in the Machine*. Hutchinson, U.K.: Macmillan, 1967.

[92] K. Arthur, "Beyond atomism and holism—The concept of the Holon," *Perspect. Biol. Med.*, vol. 13, no. 2, pp. 131–154, 1970, doi: 10.1353/pbm.1970.0023.

[93] H. A. Simon, *The Sciences of the Artificial*. Cambridge, MA, USA: MIT Press, 1969.

[94] P. Valckenaers and H. Van Brussel, *Design for the Unexpected: From Holonic Manufacturing Systems Towards a Humane Mechatronics so*. Oxford, U.K.: Butterworth-Heinemann, 2015.

[95] W. Derigent, O. Cardin, and D. Trentesaux, "Industry 4.0: Contributions of holonic manufacturing control architectures and future challenges," *J. Intell. Manuf.*, vol. 32, no. 7, pp. 1797–1818, Oct. 2021, doi: 10.1007/s10845-020-01532-x.

[96] A. Bonci, M. Pirani, and S. Longhi, "Tiny cyber-physical systems for performance improvement in the factory of the future," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1598–1608, Mar. 2019, doi: 10.1109/TII.2018.2855747.

[97] G. Casini and U. Straccia, "Defeasible RDFS via rational closure," 2020, *arXiv:2007.07573*.

[98] U. Straccia, "How much knowledge is in a knowledge base? Introducing knowledge measures," in *Proc. 24th Eur. Conf. Artificial Intell. (ECAI)*, 2020, pp. 905–912, doi: 10.3233/FAIA200182.

[99] T. Parr and K. J. Friston, "Generalised free energy and active inference," *Biol. Cybern.*, vol. 113, nos. 5–6, pp. 495–513, Dec. 2019, doi: 10.1007/s00422-019-00805-w.

[100] H. Von Foerster, "Cybernetics of epistemology," in *Understanding Understanding: Essays on Cybernetics and Cognition*. New York, NY, USA: Springer, 2003, pp. 237–241, doi: 10.1007/0-387-21722-3_9.

[101] P. E. Weston and H. V. Foerster, "Artificial intelligence and machines that understand," *Annu. Rev. Phys. Chem.*, vol. 24, no. 1, pp. 353–378, Oct. 1973, doi: 10.1146/annurev.pc.24.100173.002033.

[102] R. K. Agrawalla, "When Newton meets Heinz Von Foerster, complexity vanishes and simplicity reveals," *Kybernetes*, vol. 44, nos. 8–9, p. 1193, 2015, doi: 10.1108/K-11-2014-0265.

[103] V. Lepskiy, "Evolution of cybernetics: Philosophical and methodological analysis," *Kybernetes*, vol. 47, no. 2, pp. 249–261, Feb. 2018, doi: 10.1108/K-03-2017-0120.

[104] L. H. Kauffman, "EigenForm," *Kybernetes*, vol. 34, nos. 1–2, pp. 129–150, Jan. 2005, doi: 10.1108/03684920510575780.

[105] H. Von Foerster, "Objects: Tokens for (eigen-) behaviors," in *Understanding Understanding: Essays on Cybernetics and Cognition*. New York, NY, USA: Springer, 2003, pp. 261–271, doi: 10.1007/0-387-21722-3_11.

[106] M. Füllsack, "The circular conditions of second-order science sporadically illustrated with agent-based experiments at the roots of observation," *Constr. Found.*, vol. 10, no. 1, pp. 46–54, 2014.

[107] C. Rovelli, "The relational interpretation of quantum physics," 2021, *arXiv:2109.09170*.

[108] L. H. Kauffman, "Cybernetics, reflexivity and second-order science," *Constr. Found.*, vol. 11, no. 3, pp. 489–504, 2016.

[109] M. D. Mesarovic and Y. Takahara, *General Systems Theory: Mathematical Foundations*. New York, NY, USA: Academic, 1975.

[110] D. C. Horsman, "Abstraction/representation theory for heterotic physical computing," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 373, no. 2046, Jul. 2015, Art. no. 20140224, doi: 10.1098/rsta.2014.0224.

[111] M. Pirani, A. Bonci, A. Cervellieri, and S. Longhi, "On the synthesis of holonic management trees," in *Proc. 26th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2021, pp. 1–4, doi: 10.1109/ETFA45728.2021.9613585.

[112] E. F. Codd, *The Relational Model for Database Management: Version 2*. Reading, MA, USA: Addison-Wesley, 1981.

[113] C. J. Date and H. Darwen, *Found. for Future Database Systems: The 3rd Manifesto*. Reading, MA, USA: Addison-Wesley, 2000.

[114] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: A review of recent progress," *Rep. Prog. Phys.*, vol. 81, no. 7, Jul. 2018, Art. no. 074001, doi: 10.1088/1361-6633/aab406.

[115] R. F. Molanes, K. Amarasinghe, J. Rodriguez-Andina, and M. Manic, "Deep learning and reconfigurable platforms in the Internet of Things: Challenges and opportunities in algorithms and hardware," *IEEE Ind. Electron. Mag.*, vol. 12, no. 2, pp. 36–49, Jun. 2018, doi: 10.1109/MIE.2018.2824843.

[116] V. C. Müller and M. Hoffmann, "What is morphological computation? On how the body contributes to cognition and control," *Artif. Life*, vol. 23, no. 1, pp. 1–24, Feb. 2017, doi: 10.1162/ARTL_a_00219.

[117] G. Resconi, X. Xu, and G. Xu, "Morphogenetic computing in genetic algorithms," in *Introduction to Morphogenetic Computing*. Cham, Switzerland: Springer, 2017, pp. 107–132, doi: 10.1007/978-3-319-57615-2_9.

[118] A. Bonci, S. Longhi, and M. Pirani, "IEC 61499 device management model through the lenses of RMAS," *Proc. Comput. Sci.*, vol. 180, pp. 656–665, Jan. 2021, doi: 10.1016/j.procs.2021.01.288.

[119] A. N. Raikov and S. A. Panfilov, "Convergent decision support system with genetic algorithms and cognitive simulation," *IFAC Proc. Volumes*, vol. 46, no. 9, pp. 1108–1113, 2013, doi: 10.3182/20130619-3-RU-3018.00404.

[120] A. Bonci, P. Clini, R. Martin, M. Pirani, R. Quattrini, and A. Raikov, "Collaborative intelligence cyber-physical system for the valorization and re-use of cultural heritage," *J. Inf. Technol. Constr. (ITcon)*, vol. 23, no. 15, pp. 305–323, 2018.

[121] A. Raikov, "Manufacturer's strategic risk temperature assessment with convergent approach, cognitive modelling and blockchain technology," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1289–1294, 2019, doi: 10.1016/j.ifacol.2019.11.376.

A. N. Raikov, M. Pirani: Human-Machine Duality: What's Next in Cognitive Aspects of Artificial Intelligence?

IEEE *Access*

ALEXANDER N. RAIKOV graduated in radio physics and electronics from the Moscow Power Engineering Institute, in 1974.

He is currently a Specialty Engineer-Radio Physicist. Prior to 1993, he has developed automated control systems for the Russian Government. In 1993–1999, he was the Head of the Information-Analytical Department, Russian President Administration, and a Coordinator of the Situation Centre of the Russian President. Since 1999, he and his team has been completing about 60 projects for the Administration of the Russian President, the Ministry of Education of Russia, the Ministry of Economic Development of Russia, the Government of the City of Moscow, the Republic of Kazakhstan, BCG, and Deloitte. He has been the Doctor of technical sciences, a State Advisor of the Russian Federation, a Winner of the Russian Government Award in the field of science and technology, a Full Member of the Russian Academy of Natural Sciences, a Full Member of the Academy of Forecasting—Russian Branch of the International Academy for Future Research, a Full Member of the International Academy of Sciences of Information, Information Processes and Information Technologies, a Full Member of the International Academy of Informatization, a Professor of MIREA—Russian Technological University, the Leading Researcher of the Institute of Control Sciences of Russian Academy of Sciences, the Head of the Artificial Intelligence Department, National Centre of Digital Economy, Lomonosov Moscow State University, the Director of the Centre for Artificial Intelligence of the International Research Institute of Management Problems, and the Scientific Leader of the Expert and Analytical Centre of the Institute of Tax Management and Real Estate Economics of the Higher School of Economics. He is an Expert of the Russian Academy of Sciences, as well as many development funds and international journals. He is the author of about 500 scientific publications, nine patents, and including ten monographs in the field of strategic planning, artificial intelligence, socio-economic development, information and analytical technologies, situational centers, decision support systems, and network expertise. He has authored a book *Cognitive Semantics of Artificial Intelligence: A New Perspective* (Springer Singapore, 2021). He annually participates in foreign scientific conferences in USA, Japan, South Korea, Italy, Spain, Germany, U.K., and Brazil.

MASSIMILIANO PIRANI (Member, IEEE) received the M.S. and Ph.D. degrees in electronics and communications engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2000 and 2004, respectively.

He worked as a Research Manager and the Head of software departments in industrial and consumer electronics sectors. From 2004 to 2015, he worked as a Computer Science Researcher, a Research Manager, and the Head of software departments in industrial and consumer electronics sectors. He was mainly involved in artificial intelligence, smart control technology, networking and web technologies applied to bioprocesses, industrial systems automation, and smart home. He has been a Research and Innovation Manager and a Development Consultant in embedded electronics systems, the Industrial IoT, intelligent control, and distributed computing. He has been an Adjunct Professor on Internet, computer and computer networks, and mathematics at Marche Polytechnic University. He joined the Department of Information Engineering, Marche Polytechnic University, in 2016, where he is currently a Researcher in cybernetics, holonic multiagent systems, intelligent manufacturing systems, cyber-physical systems of systems, control theory, extended artificial intelligence, and computer science. He has participated in the research and management in four EU FP5/FP7 projects and participates in two Horizon 2020 projects contributing in the development of holonic systems and architectures for the management of building construction processes and in food supply chain. His initial research interests included identification of nonlinear systems, higher order statistics, nonlinear signal processing, audio electronics, speech processing, and loudspeaker systems enhancement.

Dr. Pirani is a member of the IEEE IES Technical Committee on Industrial Agents of the IEEE IES Technical Committee on Technology Ethics and Society. He has contributed to the developments of the IEEE Standard 2660.1^{TM} 2020: IEEE Recommended Practice for Industrial Agents.

● ● ●