Piotr (Peter) Bołtuć

PHILOSOPHY AS A THEORY OVER THEORIES[1]

ABSTRACT

We view philosophy as paradigm setting: largely, spread over leading sciences of the epoch, as well as the main developing technologies, and even socio-economic and managerial patterns. This is, obviously, a "regulatory definition," not quite a descriptive one. We examine whether it is the science of sciences, or the science over the sciences. Thus, it is not quite a meta-science. Our point is not to view philosophy as a methodology of science, or as its maid (*ancilla*). Philosophy is viewed as the pinnacle of the sciences, providing them with ontological and axiological meanings. Here is one proposed definition: Philosophy is built upon the sum of general theories of all leading sciences (broadly understood); it is a theory based on this sum. The aim of philosophy so defined is to stipulate and approximate veridical worldviews, rooted in the strongest available background, which is largely the background provided by the sciences, but not quite limited to what is scientifically provable at a given point in time—this last clause is due to temporary limitedness of any science, always existing at a given time-slice. Thus, limited dependency on any principles, not only factual statements. As we know from Albert Einstein's relativity theories and other scientific revolutions, both factual statements and higher-level principles, are always already inductively questionable, e.g., through inference to the best explanation following pragmatic, context dependent, criteria of what counts as "the best" of explanations. We also question the intuitive requirements of physicalism that are crucial to Daniel Stoljar's thesis that physicalism cannot be properly defined. In contrast to the broadly scientistic predilection beneath the approach in the main bulk of this article we also need and require a philosophical focus on the human existential condition, which is complementary to, and not contradictory

---

[1] *Footnote for Polish-reading colleagues.* "Filozofia jest sumą interesujących teoretycznie nauk i teorią tej sumy". Takiej filozofii uczyła mnie moja matka, gdy byłem dzieckiem. Wydaje mi się, że definicja ta pochodziła od jednego z jej psychologizujących profesorów filozofii, Władysława Witwickiego lub Stanisława Ossowskiego. Uważam jednak, iż zawiera ona pozytywistyczny oddźwięk, dlatego ostatnią część definicji zamieniłbym na „...i ogólną teorią opartą na tej sumie". W tym sensie odnoszę się do pardygmatyczności filozofii wobec nauk, technologii, a także szeroko-pojętych nauk społecznych, w tym teorii zarządzania.

with, the above definition of philosophy. The proposed approach may be viewed as an Enlightenment approach, aware of its strengths and limits; thus, with a post-Enlightenment zing.

1. AN UNEXPECTED INTRODUCTION

It needs to be stated that philosophy is to the other domains of knowledge like *The Song of Songs* to the other artistic, and metaphysical, texts. But this reference to the Sacred Writings would derail the meaning thatI want to convey.This is because in my native language, Polish, *The Song of Songs* is oft translated as "*The Song over the Songs*" (*Pieśń nad Pieśniami*). This is hard to grasp even in Yiddish, where *Song over Songs* would translate as "*lid iber lider*." Yet, "*lid iber lider*" has a more natural translation as the "song about songs." The title would be translated from Yiddish into German, as "*Lied über Lieder*," which is also the literal translation of the title in Polish. However, in German the book is named *Das Lied der Lieder*, which translates into "*The Song of Songs*." The above set of conceptual intricacies may partly explain some attachments to the view of philosophy different from the one I am here to present.

The above point is not to draw close analogies or direct causal connections between *The Song of Songs* and the philosophy of science. It rather illustrates how little distinguishes the concept of internal analysis (inter-paradigmatic) and paradigmatic change with *über-positioning* the new paradigm. Of course, Hegelian dialectics applies to paradigmatic change, where victorious antithesis morphs with "thesis," into a synthesis, tough on its, victorious, account. This is relevant since early readers of various versions of this paper tended to be stuck on the level of the Popperian or other intra-paradigmatic, methodological models. We focus on philosophy in the current world-period, which is that of a series of rapid paradigmatic changes. This is due to the fact that philosophies of the last century did not really incorporate most of the paradigmatic changes in the sciences of Einstein and Heisenberg, despite radical attempts to do so early on.

*

I view the change between GOFAI (Good Old Fashioned Artificial Intelligence) and AGI (Artificial General Intelligence) as deeply paradigmatic. Philosophy today must lead the way, in the set of changes that incorporate not only the latter revolution, but also the consequences it brings to the theory and practice of Digital Transformation in business and management. We also need to follow Roger Penrose and others in re-evaluating and incorpo-

rating in our worldview the philosophical gist of quantum physics. This is essential in the phase of quantum computing as well as the quantum explanations of human and other animal brains.

For many persons philosophy plays the conceptual role of *The Song of Songs*, interpreted as the song *about* songs; as such, it is the tool of methodological analysis of the other domains. The followers of this view are stuck with merely positivist take on philosophy as the general methodology of the sciences. However, the new paradigmatic change seems to come from technology, as the 4th and 5th industrial revolutions extend to the social, economic, and political practice and thought. For me it is the theory above and beyond the most general theories of the sciences and other domains—a theory that presents transdisciplinary paradigms—and much less so, an inter-paradigmatic (limited to the *normal science periods*) chiseling of methodological and formal details.

We view these supposedly general conditions as extrapolations of "the ways of the craft" unjustifiably extrapolated. For instance, Ockham's Razor is a heuristic principle of "small brains," while big-data-based AI, may count as mid-size-brains, with its different heuristics and appropriate methodologies (Goertzel, 2006). While this sounds a bit like David Hume's skepticism, we are open to much richer ontologies informed by a new leading praxis, such Penrose's take on quantum physics, or Ben Goertzel's on Artificial General Intelligence.

2. DEFINITIONS

The primary interest guiding this paper is in "philosophy as the science over the sciences." It is not quite a meta-science since the point is not to view philosophy as a methodology of science, or—to use a Medieval expression in a different context—*as a maid to the sciences*. Philosophy, in inter-paradigmatic change, is viewed as the pinnacle of the sciences (Boltuc, 2022; 2023), able to give ontological and axiological meanings to their findings, set up in a moderately unified perspective of the intellectual search for a new paradigm..

2.1. Definitional Objectives

I pose that philosophy should provide us with the image of the world informed by natural sciences; yet, reaching beyond their scope to the realm where competencies of specific sciences fade away, leaving the elbowroom for systemic theories broader than any of the sciences.

Such a meta-science takes into account multidisciplinary or rather trans-disciplinary investigations on patterns and regularities across various disci-

plinary methodologies and the local scope of questions any particular science seems tasked to address. Mathematics does so at the most abstract formal end of the spectrum, while philosophy belongs to the other end of the spherical spectrum, where the sciences become *nearly* a member of the humanities but not quite so. Philosophy meets mathematics, through formal methods, including but not limited to formal logic, but differs from basic-level mathematics by building structures and conceptual Gestalts that transcend given formalizations.

The humanities and the sciences *seem not* to lie on the same spectrum. Yet, the humanities cross the spectrum of the sciences at the intersection with philosophy. Thus, philosophy of the kind we are talking of, is primarily a meta-science, or rather the most theoretical of the sciences, which only secondarily belongs to meta-humanities (or, as I prefer to say) to transhumanities (Boltuc 2022). Those spectra interlink at the issues related to the human condition—individual and social, as well as our bio-medical, psychological, axiological, and goal-related (evolutionarily given and bound) and epistemic ramifications.

The intra-paradigmatic aspect, viewed as methodology is helpful as a tentative working sub-paradigm within the realm of a stable phase in the sciences. They should not be viewed as universal tools. Even logic encounters its practical limitations, with an occasional room left for the paraconsistent logics (Goertzel, 2022; Boltuc, 2023) including quantum logic as well as non-axiomatic logics (Pei Wang).

Many people get perplexed by the scientific revolutions, and such discomfort leads to their tendency to dismiss the tentative character of stable-stage methodologies. On the contrary, other people find the stable, inter-paradigmatic phase as boringly tedious and dangerously dogmatic. The author is closer to the latter group.

To sum up, I view philosophy—or the philosophies—as primarily located at the extended spectrum of the sciences. Their image of the world approaches some level of objectivity thanks to scientific, and these days also engineering, methods. Objectivity in the humanities is more tangential, dependent on socio-psychological visions, projections, and loosely verifiable ideations.

2.2. Defining Philosophy as a Metaparadigm

We begin with Definition 1, which I inherited, in a rather straightforward sense. This is followed by Definition 2, which fixes what, in Definition 1, could be interpreted as its twist towards the methodological positivism of the mid 20th century.

Definition 1

Philosophy is the sum of general theories of all theoretically interesting sciences, and a theory of this sum. This is what my mother told me when I was five or six. I have no idea where this definition came from (though I tried to locate the source). I grew to like and take it as my starting point, kind of. However, I am not sure I quite understand the final clause of this definition.

What I understand is its first part; I know what the total sum of the general theories in all the main sciences would be. The "theoretically interesting" clause seems to exclude the theories of some of the more applied sciences whose general theoretical import may be of little, if any, paradigm building value. We may put together the remaining scientific theories, e.g., by printing them in the same volume. Yet, I doubt that such a sum of theories would create a helpful set—this, however, is a worry for further consideration.

Here is the worry at hand: While I can see—in general terms—what such a sum of general theories would be, it is not clear how it would be considered philosophy. What seems to be meant in the first part of Definition 1 is this: All general sciences, separately, or in groups, could answer some questions that reach to the limits of—or even just beyond—the scope of those sciences. I am not sure how to understand the second part of Definition 1: What is meant by philosophy as "a theory of the sum of scientific theories"? It could be meant to be a meta-reflection on those theories, pertaining to their methodology, or perhaps to their structure (including the topic whether their sum is coherent, consistent, or whether they follow some common patterns). Those are largely methodological, considerations. Yet, the above clause could also mean something even more ambitious. In the latter instance, the gist of philosophy would be defined as a positive higher-level science, built as a theory based on the main general sciences—which would amount to a scientifically-based worldview. Such philosophy clearly belongs amongst the sciences. But who would be qualified to become a philosopher equal to the task of creating such philosophy, as a science at a higher level of generality? It would rather be suited for the most theoretically astute and philosophically inclined scientists or for the teams of scientists and philosophers. I find this approach attractive and ambitious, but it is not the most natural reading of Definition 1, which seems to be geared towards ohilosophy as methodology of the sciences. To express something like the latter project more clearly, we reformulate definition 1.

Definition 2

Philosophy is the sum of general theories of all theoretically interesting sciences, *and a theory built upon this sum*.

This theory is not *abou*t the sciences, in terms of their methodology, or broader meta-analysis. It is a theory above scientific theories. Thus, my di-

vagations about understanding of the "*Song above/über songs* or *The Song of Songs* are structurally relevant for the main point of this paper. We view philosophy as a theory "*above*-disciplinary-theories," even above interdisciplinary ones. As a general theory over the sciences philosophy is the roof that crowns disciplinary, or multi-disciplinary, walls and other elements of paradigm as a building-construction.

It should be clearly understood, that Definition 2 does not pertain to every kind of legitimate philosophy. It refers to the one appropriate and highly needed, meaning of philosophy as a theory that that notices and implements the paradigmatic wholeness[2] to the products of the sciences and other relevant domains (e.g. engineering, or socio-economic structure).

For instance, *De revolutionibus orbium coelestium* by Copernicus is a book in astronomy with such a philosophical—paradigm changing[3]—dimension and intention. All the five examples of scientific revolutions essential to pruning the human self-image, discussed by Luciano Floridi (Copernican, Darwinian, Freudian, and Turing's), are the examples of such a philosophical impact on the scientific theories. They can be extended to Newtonian, Einsteinian and Heisenberg's revolutions in physics, and related sciences. It is worth observing, that such revolutions are hardly ever initialized by professional philosophers, rather the scientists. They may come from domains other than natural sciences but come from economics[4]) and other domains. Even philosophers such as René Descartes, whose thought formed a new paradigm in thinking, came not only from Descartes as a philosopher, but also from his genius as a mathematician. My approach comes from largely from Artificial General Intelligence (AGI), as presented by Ben Goertzel, Stephen Thaler, and, somewhat tangentially but importantly, Aaron Sloman (2020).

3. PHILOSOPHY AS A *THEORY over THEORIES* OF AGI?

The point is not that every work of contemporary philosophy has, or needs to have, anything to do with Artificial Intelligence. The main point is to present philosophy that is not philosophy of "computing" or "or AI" but rather philosophy as an extension of theories withing AI, especially of those that aim at getting closer to Artificial General Intelligence.

Here is the actual claim: Artificial Intelligence is becoming, rapidly, the area of engineering most influential on our scientific paradigm, and even as a way of life. Yet, only AI on its way to AGI (Artificial General Intelligence),

---

[2] Transdisciplinary explicability; not limited to local domains.

[3] Paradigmatic wholeness is likely enhanced by paradigmatic revolutions, which terminates dominance of pedantic, local conventions of meaning.

[4] Including Adam Smith; also Karl Marx

is a truly disruptive field. Using Floridi's approach, only AGI changes our world to the extent comparable to that of Copernicus, or Darwin.

3.1. Alan Turing as the Harbinger

Contrary to Luciano Floridi, Jack Copeland, and also to my earlier works, the set of disruptive changes we start encountering, is not Alan Turing's revolution, though Turing's work and forward-thinking ideas paved the way for it.

This approach is similar with my take on the Freudian revolution, which, although originating from Sigmund Freud's overemphasized ideas on subconscious role of sexuality, became truly disruptive only with the works of Benjamin Libet (questioning the possibility of conscious decisions, based on the fact that we often act overly fast to have had conscious access to the issue at hand), especially in its interpretation by Max Velmans. The issue of Freudian revolution seems much more than a conundrum, only if we take into account the explanation by Roger Penrose, of human consciousness' jump to quantum space, as explanation of acting in different time-space. This topic is beyond the scope of this paper, though the issue is of high philosophical import.

It seems that AGI (or, AI on its way to AGI) is the gist and climax of digital revolution, as far as we can tell based on what we observe (referring to epistemological attitude of Phei Wang methodology, which is central to NARS). Having engaged in, somewhat shallow, interpretation section of the Torah, pertaining to *The Song of the Songs*, let me mention a Christian analogy from the Gospels. St. John the Baptist, as the harbinger of Jesus is not supposed to be the father of Christianity. Judaism, Christianity and Muslim religions all have their prophets that consist not the center, but the entry point of the crucial personality. In the same way, Turing as the harbinger of digital revolution, is not quite its father, or pinnacle. I understand that this is a non-standard view and several of my good friends, and colleagues, will take it as a philosophical blasphemy. The centrality of Turing's work is looming large still in the perspective dominated by advanced version of ChatGPT and other low-scope, risk related AI—most of which can be recognized by facing the catastrophic failure upon attempts at re-training. From the viewpoint of AI leaning towards AGI (AI2AGI_Turing is the harbinger of AGI revolution, while its apex is still to come, in a short time—period. Well, this sounds semi-religious, but every look beyond the horizon, even short of self-assured futurology, has the zing of prophecy, while becoming reality "as we go," often in rather short order.

The truly disruptive movers and shakers come when the environment is ripe. Nicolaus Copernicus was preceded by Aristarchus and his ancient followers, also being aware of criticisms of the geocentric Ptolemaic system by

German astronomers of his times, as well as by Jean Baptiste de Lamarck, Pierre Louis Maupertuis and many others.

To use a different instance, the point is not that Charles Darwin's mechanical understanding of evolution was always already the best version. Under some reformulation some aspects of Lamarck's approach may perhaps need to be revisited. Yet, Darwin was the mover and shaker who brought evolution to the forefront of the worldview of nearly all enlightenment societies

To some up this section: Arguably, Freud was one of the harbingers of subconsciousness, which currently is so important not only in neuroscience but even in AI (Troy Kelley 2014). Analogously, if Turing is the harbinger, and the prophet, I venture to pose that the center of digital revolution is AGI.

4. AGI AS A CHANGE AGENT

I envisage a reviewer pointing out a grammar error in the above title. Human beings that initiate changes, especially on someone else's biddings, politicians, business leaders, spies, etc. are human designers and starters of the causal chain leading to a major change, not their tools and designs (such as robots), embedded or otherwise. Well, this claim is no longer obvious.

AGI pioneers, such as Ben Goertzel and the teams working with and around him, are the towering figures. So is Pei Wang and many others. However, there is no longer an overwhelming need to find a human beyond the machine. I shall tackle the last point with examples.

4.1. AlphaGo and AlphaZero: the paradigmatic lessons

The AlphaGo story from 2017 sounds like the old news. The game of Go is so complex that it was viewed that, unlike chess, it would never by mustered by AI at the level of human champions. However, AlphaGo, has been trained by human masters and programmers, which resulted in it winning consistently over human champions.

However, not everybody knows the AlphaZero story. Even fewer seem to draw from it a paradigmatic lesson. The story about it in the journal *Nature* is prefaced with the sentence "A long-standing goal of artificial intelligence is an algorithm that learns, tabula rasa, superhuman proficiency in challenging domains" (Silver, 2017). AlphaGo Zero trained on its own, "based solely on reinforcement learning, without human data, guidance or domain knowledge beyond game rules" (Silver, 2017). After 8 hours training it met and surpassed the skills of the human-trained Alpha Go.

This point teaches that, in a short run, human-trained robots turn out inferior to those self-trained—at least in complex strategic games. This point is

vital. Since every complex activity, such as logistics, engineering, military and political jostling, even discovery, can in principle be transcribed as a series of strategic games—this implies that AI not trained by humans is better in those strategic games or groups thereof.

Importantly, this runs counter to the hurrah optimism that AIs will be merely the solvers of formal problems; yet, in real-life applications they would remain merely human helpers. In terms of strategic games humans would hardly every be equal or superior players over the strong AGIs. There is one set of exceptions to this: Humans are still quite better than AI in understanding complex situations, social, sometimes even physical interactions. To explain this point, one is behoved to revisit the work of Miriam Yevick who argues in favor of complementarity between formal knowledge (what can be described in predicative language and its logic) and what she calls "holographic" epistemic level, where we interact directly with the objects. The formal language is good enough for expressing, sketch out, issues within general knowledge. However, only holographic language can convey the depth of details efficiently. This is a more advanced formal take on the distinctions between knowledge by description and knowledge by acquaintance (Benzon, 2022).

Today's advanced cognitive machines, such as DABUS (Thaler, 2021) are crossing the line between being just mathematical-logical engines, the way GOFAI functioned and AGIs. This is in part thanks to incorporating various sensors, both by robots and AI running on the computers that are linked to the computer-sensors (including e.g. ophacotry sensors), or the sensors (e.g. based on the sense of vision) that get knowledge from the web.

This is the kind of philosophical lesson that seems to be nuanced, yet provides a decisive proof in favor of AGI and its potential for development. Several people shy away from "counterintuitive" philosophical conclusions. But what is intuitive changes every half-generation, if not more often.[6] Such *connundra* pave the way for paraconsistent vision of the world, even paraconsistent ontologies. This leads us to the next philosophical problem.

4.2. Paraconsistent logic for AGI: systemic incompatibilities in Humans and AI

Ben Goertzel argues that human actions, practices and even ethical values are paraconsistent. In this he follows the works of Zach Weber (Weber, 2007). Thus, a humanoid companion or a robot meant to cooperate with humans smoothly (instead of patronizingly) would need to follow largely some form of paraconsistent logic (Goertzel, 2021a; 2021b; 2021c).

---

[6] John Barker (2008,2009) has helped me sharpen this kind of approach, based upon his formal presentation of the liar's paradox as a *conundrum*, which we need to live with, not dismiss off hand.

Such paraconsistency is easily explained by research of Joshua D. Greene and Jonathan Heidt in empirical moral psychology. Human beings have at least five "moral senses," which Heidt calls moral foundations. Those are: (1) harm/care, (2) fairness/reciprocity, (3) ingrouployalty, (4) authority/ respect, (5) purity/sanctity. Later, he added (6) *libertyversus*enslavement. All these categories are based on psychological research. The first two are common to all "normally" developed human beings—as well as several mammal species such as apes, dogs, horses, rats etc. Interestingly, those moral senses are placed in various (evolutionarily different) location in the brain. While human mind has ways and means to build—always *ad hoc*— compromises among those six sets of values, there is no overarching value or classification. In ethical theory, this is consistent with over a generation older theories by William D. Ross' based on *prima facie* moral reasons (agent-relative objective moral reasons) and Jonathan Dancy (moral partic- ularism) (Boltuc, 2023).

Goertzel's contribution is to extend the use of paraconsistent logic used by Weber largely to human ethics, to the domain of human actions and hier- archy of goals, while then applying this approach to AI.

Functional paraconsistencies in human behavior account for structural complexities Goertzel argues that they would be eliminable in creatures a level of magnitude smarter than us. I daresay, that those creatures would be even more flexible and adaptable to multiplicity of goals and values with context-dependent value-structures, so that paraconsistencies would be even more subtle and nearly omnipresent..

Goertzel's formal arguments are complex, and go beyond the scope of the current article; they are presented in more detail in (Goertzel, 2021). Here I can share just two crude, intuitive accounts. The first one comes from Goertzel's work, the other one also comes from Goertzel, but rather through a deviant causal chain.[7]

The first simple example shows that, we may focus on the trivial feature of paraconsistent logic, which is captured by propaedeutic version of "logic of times." If we want to account for the state of normally functioning traffic lights, we need to introduce the time factor in their longitudinal description, to make it into something like: time10.00 green; time 10.01 orange; time 10.01,20 sec red, 10.02- red and so on. This operation may be described as paraconsistent due to the fact of veridical statements depending on the time, as a changeable factor.

The second simple example crossed my mind when attending, via Zoom, one of the recent AGI conferences organized and attended by some of the top world-experts in AI, AGI, and computer science. Due to the use of a nov- el, highly complex, portable microphone, highly awaited presentation by

_____

[7] The issue of deviant causal chains is big in philosophy of law and British analytical ethics.

Josha Bach was delayed by about 20 minutes. Interestingly, after the lunch break the same problem re-appeared and took even longer to solve. The problem was the lack of interoperability among the systems, probably primarily by the "super microphone." In this context, paraconsistency of the communication space among complex, mutually not translatable systems or languages comes from high communication criteria that go beyond what's needed for the occasion, sort of overly high granularity. In this context, paraconsistency of mutual medium of communication is a way to lower such granularity, to allow 'good enough for the occasion' mutual communicative interaction.

There is a "deviant causal chain" (Peacocke 1979) between Goertzel (*qua* a conference organizer) and the above account. Actually, while the "easy-peasy" argument put forth above seems mine, its deeper formal explanation comes from Goertzel's, formalized account of "paraconsistent interzones," which predates the abovementioned event.

Briefly speaking, "A paraconsistent interzone is a space in which limited forms of paradox and inconsistency are permitted to flourish, and which serves in part as a medium for interoperation and conversion between other spaces that are, in themselves, more narrowly consistent;" for details consult (Goertzel, 2021).

4.3. Brute force speed with maneuverability

In last years it came to public attention that India (partly based on Russian technology) has developed hypersonic (Mach 2.8) and highly maneuverable missiles BrahMos with range of about 500 km. Those are capable of carrying nuclear weapons.[8] Indian and Russian propaganda claim that those missiles are not vulnerable for any currently operational air-defense systems. However, dominant view is that "Speed of Mach 2.8 achieved by the missile pose no difficulty for S-400 operators, they will get more than 70 seconds to react against the incoming threat, which is very normal."[9]

However, situations with 7 seconds for reaction, or less, may happen, especially in the context of "surprise" attacks at non-standard war theatre. In those cases, human involvement becomes cumbersome, for the most part detrimental, due to the slow speed of human reaction and decision times. In such instances, advocating that humans must always be in charge of lethal weapons becomes inconsistent with realty and therefore highly unhelpful, even within the just war doctrine.

---

[8] P. Satam, The EurAsian Times, September 11, 2023; The article disappeared during my reading, circa 10.00 am Central European Time, published and read on the same day. Then the paper disappeared perhaps due to editorial changes or some kind of censorship; checked again September 23, 2023.

[9] J. Leci, *Can S400 Shoot Down a Brahmos Missile?*, Quora; https://www.quora.com/Can-S400-shoot-down-a-brahmos-missile#;" retrieved September 11, 2023.

AI taking over passenger jet landing process, as well as complex hearth and brain surgical procedures, and even car driving under extreme conditions—is not only an option, but a moral necessity. This is due to disproportionately high chances of success, compared to such chances by well-trained human agents (pilots, drivers or surgeons).

4.4. AI in the courts of law

Social discrimination, or acceptance, of advanced AIs is still a volatile topic. Those educated in environmental justice, thus familiar with the notion of human chauvinism, may, or may not fully grasp this problem, since ecology is sometimes based on emotional defense of living things, as the only bearers of consciousness.[10] But chauvinism is never a good solution for complex relations with intelligent others.

Some people tended to argue in favour of the tactics of oppression during the colonial and post-colonial era, when any attempts at autonomy by the oppressed resulted in even more oppression. Sorry to say, but Isaac Asimov's robot ethics, relying on strict servitude of robots and never their recognition as rational beings was in principle racist.

The lack of recognition of other intelligent beings would be highly morally objectionable to Kant, for whom ethics pertains to all "intelligent beings." Yet, for Kant intelligent beings are those active "transcendental subjects" that put their vague apperceptions in the categories of time and space, but also of teleological and ethical categories. After Einstein's relativity theories we can no longer view time and space as absolute, unchangeable features of our mind. The same goes for ethics, which is now viewed as more and more contextual, even paraconsistent[11]. Thus, Kant's trust in a priori forms and absolute dimensions grasped by intelligent minds seems overly formalistic, and dry of empirical context.

For AIs to partake in the real world, not solely in the world of formalisms, always already sketchy and devoid of empirical content, they have to experience, what Yevick calls holographic consciousness. Today, we would think of it as of the sensors that accompany AIs cognitive mechanisms. Knowledge by acquaintance is not quite reducible to knowledge how, or knowledge by description. This is not just true of embedded AIs in robotics. We talk about sensory perceptrons visual, audible, but even olfactory. In fact, AI can easily process "sensory inputs" far more subtle than human beings, or other animals.

Thaler's controversial identification with his AI, cognitive engine working within the sensory space, which seems to be on its way to AGI, is an im-

---

[10] Xinyan Zhang in yet unpublished paper *The Ontological Essence of Consciousness,* seems to be making some such claims.
[11] See the section 4.2 qbove.

portant step in the right direction. Epistemically, the biological capabilities of one's body are relevant to one's mind only as long as the nervous system conveys them to their cognitive system, such as a brain in most of the intelligent animals. Generators of perceptions work best if they are in fact generated by empirical/objects. I discussed this problem in philosophy of mind, as the issue of the amount of unambiguous information on one page full of legible words versus a page of a realistic picture, or painting. This may have come, in part, from my conversations with Miriam Yevick at Princeton University, when I was a graduate student. But the gist is multifarious dimensions, nicely grasped by Rachel St. Clair at Simuli Team, including her work on hyper-vector analysis (St. Clair, 2023; Goertzel, St. Clair, 2022).

To bring in another aspect of "AI in the course of law." A few months ago, several colleagues of mine and I myself, received enquiries from the office of a notable US politician asking to present the drawbacks that come from advanced AI. When I volunteered to present pros- and cons- the offer went dry. Many politicians in the EU and international organizations around the world seem to follow a similar agenda.

Such culture of fear is also propagated by the courts. Even self-professed progressives are often weak in the area of sciences and easily scared by technological news. Lack of competence leads to procrastination, and trying to stay away from from what is viewed as the hot potatoes AI technology throws their way. This approach was visible with Thaler's attempts to get the patents to DABUS, his imagitron active in semi-autonompus research and discovery engineering. Thaler argues that a human expert given the same, minimal, amount of guidance would be judged as a patent owner. Thus, it would be misappropriation of a patent by Thaler, the company's owner, to claim patent ownership.

Except for the Australian justice Beech and a few other lone voices around the globe, court majorities harkened back on the blunt philosophical platitudes, referring to the human dignity and similar vaguely explained ideas. While Thaler's legal show may have been premature, it demonstrates how far prevalent law (and politics) is from integration of AI in the society.

Why is such integration a good thing. Here come a few examples (some already mentioned in the currant paper):

It is hardly controversial that AI driven cars are at least 6 times safer to drive, for those inside and outside the vehicle than those driven by average quality human drivers. But the courts in most jurisdictions draw their legs and focus on sophomoric theoretical questions of who is legally responsible for those few accidents that happen nevertheless. One could propose that this situation is similar to those, where a human driver is not legally competent (e.g. went crazy) or dead. The people in the car producing, and/or selling and/or using, and among those in the traffic, who created the conditions

that made the accident likely may be liable, both fiscally and criminally. Thus, the point is not to put AIs at the helmet of a corporation or any other legal entity—not for a while. This is due to two reasons:

A. AIs are too weak in their grasp of the social context.

B. The society needs a gradual shift towards AI, or AGI, being in charge of day-to-day operations. This may originate from emergency situations, such as speedy war theatre or economic warfare (e.g. a split-second investment theatre).

AIs functioning in the social environment, is however, the sole way for intermingling, as well as meaningful cooperation between human and AI intelligent agents.

CONCLUSION

How is the above discussion of near-future's AGI relevant to the article on the need to expand the role of philosophy? First, philosophy (Floridi, 2014) as well as philosophical/paradigmatic aspects of management theory (BCG, 2015) demonstrate how paradigm-changes happen every few generations though they tend to be misinterpreted as calamities both by the conservatives (Inquisition) and socialists (the luddites supported by Pierre-Joseph Proudhon). I sort of follow Karl Marx's critique of Proudhon in his *The Poverty of Philosophy*. This is for good reasons: First, Marx condemned somewhat weak philosophers, such as Proudhon, with critiques that failed to touch the real philosophers capable to thrive in the inter-paradigmatic periods. This is true in epistemology (for instance, for both Descartes and Princess Elisabeth of Bohemia, his sharp and underrated critic) but also in politics, and economics. Secondly, Marx being a brilliant critic, became an idealist by putting wishful thinking based on ideological superstructure (namely the idea of communism), in front of economic and technological sobriety. Copernicus, Darwin, even Freud, and Turing, have been inter-paradigmatic philosophers, not solely scientists. In the last instance, of Turing, I would not be surprised if the pinnacle of digital revolution was not Turing, or Goertzel, or AlphaZero but the first true AGI. Well, then it would have to be competent in paradigmatic Philosophical strategies, as any of the top philosophers in the periods of disruptive technology.

.

REFERENCES

Barker, J., *Undeniably Paradoxical*, Polish Journal of Philosophy, 2008, 2 (1), pp. 137–142.
_____ , *Disquotation, Conditionals, and the Liar*, Polish Journal of Philosophy, 2009, 3 (1), pp. 5–21.

Benzon, B., *Miriam Yevick on Why Both Symbols and Networks Are Necessary for Artificial Minds*, Lesswrong.com 6th June 2022; https://www.lesswrong.com/posts/nupJLBe2xxb B8KJxc/miriam-yevick-on-why-both-symbols-and-networks-are-necessary

Block, N., *On a Confusion about a Function of Consciousness*, Brain and Behavioral Sciences, 1995, 18 (2), pp. 227—247.

Boltuc, P., *Moral Space for Paraconsistent AGI*, in: Goertzel, B., Iklé, M, Ponomaryov, D., Potapov, D. (eds.), Artificial General Intelligence, Chapter 16, series: Lecture Notes in Computer Science, Springer Science, Switzerland, 2023 pp. 83–97.

_____ , *Transhumanities as the Pinnacle and a Bridge*, Humanities MDPI https://www.mdpi.com/2076-0787/11/1/27 vol 11/1 (2-2022)

_____ , *Strong Semantic Computing*, Procedia Computer Science, 123, February, 2018, pp. 98–103; https://www.sciencedirect.com/science/article/pii/S1877050918300176

_____ , *Church-Turing Lovers*, in: Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence; Lin, P., Abney, K., Jenkins, R. (eds.), Oxford University Press, Oxford, UK 2017; pp. 214–228.

_____ , *The Engineering Thesis in Machine Consciousness*, Techné: Research in Philosophy and Technology, 16 (2), Spring 2012, pp. 187–207.

_____ , *What Is the Difference between Your Friend and a Church Turing Lover*, in: The Computational Turn: Past, Presents and Futures? C. Ess, Hagengruber, R. (eds.), Aarchus University 2011, pp. 37–40.

_____ , *Qualia, Robots and Complementarity of Subject and Object*, World Congress of Philosophy, Boston 1998; http://www.bu.edu/wcp/Papers/Mind/MindBolt.htm

Princess Elisabeth of Bohemia, René Descartes, *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*, Lisa Shapiro (ed., trans.), University of Chicago Pres, Chicago 2007.

Goertzel, B., *Paraconsistent Interzones (and Associated Wild Speculations)*, Eurykosmotron, August 13, 2021; https://bengoertzel.substack.com/p/paraconsistent-interzones

_____ , *Exploring Open-Ended Intelligence Using Patternist Pilosophy*, IS4SI, Philosophy and Computing September 14, 2021b; https://www.youtube.com/channel/UCQ3w2Jpi6DQ f9aK51A0LLFA

_____ , *Paraconsistent Foundations for Probabilistic Reasoning, Programming and Concept Formation*, ArXiv abs/2012.14474, 2020

Kelley, T. D., *Robotic Dreams: A Computational Justification for the Post-Hoc Processing of Episodic Memories*, International Journal of Machine Consciousness, 06 (02), December 2014, pp. 109–123.

Peacocke, Ch., *Deviant Causal Chains*, Midwest Studies in Philosophy, 1979, 4 (1), pp. 123–155.

Penrose, R., *Quantum Theory & Free Will*, The Institute of Art and Ideas, 2019; https://www.facebook.com/watch/?v=428473478079787; last checked 9/20/2023.

Silver, D. et. al., *Mastering the Game of Go without Human Knowledge*, Nature, 2017, 19 October; https://www.nature.com/articles/nature24270; accessed 11 September 2023.

Sloman, A., *Varieties of Evolved Forms of Consciousness, Including Mathematical Consciousness*, Entropy, 2020, 22 (6), 615; https://doi.org/10.3390/e22060615

StClair, R., *Leveraging Conscious and Nonconscious Learning for Efficient AI*, Frontiers in Computational Neuroscience, 2023, 17, 23 March; https://doi.org/10.3389/fncom.2023.1090126; also: https://www.frontiersin.org/articles/10.3389/fncom.2023.1090126/full; accesed Sept 2. 2023.

Turing, A. M., *Computing Machinery and Intelligence*, Mind, 1950, 59 (236).

Velmans, M., *Preconscious Free Will*, Journal of Consciousness Studies, 2003, 10 (12), pp. 42–61.

Weber, Z., *On Paraconsistent Ethics*, South African Journal of Philosophy, 2007, 26 (2), pp. 239–244; DOI: 10.4314/sajpem.v26i2.31477

ABOUT THE AUTHOR – Professor, Double Doctor, Bowling Green State University (Applied Ethics); The University of Warsaw (Philosophy of Person). Professor at the University of Illinois at Springfield, USA (Department of Mathematical

Sciences and Philosophy; Computer Science); The Warsaw School of Economics, Professor SGH (Management Theory). Specializing in Digital Transformation and AGI revolution. Advisor to Simuli, an AGI startup company; https://www. simu-li.ai/team/

Email: pboltu@sgh.waw.pl